Comparative Policy Studies

Conceptual and Methodological Challenges

Edited by Isabelle Engeli
and Christine Rothmayr Allison

ecpr

Comparative Policy Studies

Research Methods Series

General Editors: **Bernhard Kittel**, Professor and Head of Department, Department of Economic Sociology, University of Vienna, **Carl von Ossietzky** Universität Oldenburg, Germany and **Benoît Rihoux**, Professor of Political Science, Université catholique de Louvain (UCL), Belgium

In association with the European Consortium for Political Research (ECPR), Palgrave Macmillan is delighted to announce the launch of a new book series dedicated to producing cutting-edge titles in Research Methods. While political science currently tends to import methods developed in neighbouring disciplines, the series contributes to developing a methodological apparatus focusing on those methods which are appropriate in dealing with the specific research problems of the discipline.

The series provides students and scholars with state-of-the-art scholarship on methodology, methods and techniques. It comprises innovative and intellectually rigorous monographs and edited collections which bridge schools of thought and cross the boundaries of conventional approaches. The series covers both empirical-analytical and interpretive approaches, micro and macro studies, and quantitative and qualitative methods.

Titles include:

Joachim Blatter and Markus Haverland
DESIGNING CASE STUDIES
Explanatory Approaches in Small-N Research

Alexander Bogner, Beate Littig and Wolfgang Menz (*editors*)
INTERVIEWING EXPERTS

Michael Bruter and Martin Lodge (editors)
POLITICAL SCIENCE RESEARCH METHODS IN ACTION

Isabelle Engeli and Christine Rothmayr Allison
COMPARATIVE POLICY STUDIES
Conceptual and Methodological Challenges

Bernhard Kittel, Wolfgang J. Luhan and Rebecca B. Morton (*editors*)
EXPERIMENTAL POLITICAL SCIENCE
Principles and Practices

Audie Klotz and Deepa Prakash (*editors*)
QUALITATIVE METHODS IN INTERNATIONAL RELATIONS
A Pluralist Guide

Lane Kenworthy and Alexander Hicks (*editors*)
METHOD AND SUBSTANCE IN MACROCOMPARATIVE ANALYSIS

Ingo Rohlfing
CASE STUDIES AND CAUSAL INFERENCE
An Integrative Framework

Comparative Policy Studies

Conceptual and Methodological Challenges

Edited by

Isabelle Engeli
University of Ottawa, Canada

Christine Rothmayr Allison
University of Montreal, Canada

palgrave
macmillan

Contents

List of Tables and Figures

Tables

Figures

Notes on the Contributors

John S. Ahlquist is Associate Professor of political science, Trice Family faculty scholar, and H. I. Romnes Fellow at the University of Wisconsin, Madison. He is also a research affiliate in political economy at the United States Studies Centre at the University of Sydney. He maintains active research interests in political economy and methodology. His work has appeared in the *American Journal of Political Science, American Political Science Review, Journal of Politics, Political Analysis* and other outlets. He has recently published *In the Interests of Others: Leaders, Governance, and Political Activism in Membership Organizations* (Princeton University Press 2013).

Sophie Biesenbender is a postdoctoral researcher at the Institute for Research Information and Quality Assurance, Berlin, Germany. Her research focuses on comparative public policy as well as policy change and international policy diffusion in environmental and social politics. She specializes in quantitative methods and applied statistics. She has published articles in *Journal of European Public Policy*, *European Integration online Papers* and *Global Environmental Change*.

Joachim Blatter is Professor of Political Science at the University of Lucerne (Switzerland). He received his PhD from the Martin Luther University Halle-Wittenberg and wrote his Habilitation at the University of Konstanz (Germany). He held positions at the University of St. Gallen, Free University of Berlin and Erasmus University Rotterdam and was visiting scholar at Harvard, European University Institute in Florence and Australian National University in Canberra. He has published extensively on cross-border institution building in Europe and North America, on environmental politics as well as on transformations of governance, citizenship and democracy. His works have appeared, among others, in *Citizenship Studies, Democratization, European Journal of International Relations, Governance, Global Environmental Politics, International Journal of Urban and Regional Research, Publius* and *West European Politics.* He has written textbooks and articles on qualitative and case-study methods – for example, *Designing Case Studies: Explanatory Approaches in Small-N Research* (2012, with Markus Haverland); *Qualitative Politikanalyse* (2007, with

Frank Janning and Claudius Wagemann) and 'In Search of Co-variance, Causal Mechanisms or Congruence? Towards a Plural Understanding of Case Studies', *Swiss Political Science Review* 14, 2, 315–56 (2008, with Till Blume).

Christian Breunig is Professor of Comparative Politics at the University of Konstanz, Germany. His research concentrates on public policy in advanced democracies, comparative political economy and political methodology. His works appear, among others, in the *American Journal of Political Science, Comparative Political Studies, Governance, Policy Studies Journal* and *Political Analysis*. He directs the German Policy Agendas project.

Ben Cashore is Professor of Environmental Governance and Political Science and Director of the Governance, Environment, and Markets Initiative at Yale (GEM). He is also the Director of the Program on Forest Policy and Governance. Professor Cashore's research interests include the emergence of non-state, market-driven environmental governance; the impact of globalization, internationalization and transnational networks on domestic policy choices; comparative environmental and forest policy development; and firm-level 'beyond compliance' sustainability initiatives. His book, *Governing Through Markets: Forest Certification and the Emergence of Non-state Authority* (with Graeme Auld and Deanna Newsom), was awarded the International Studies Association's 2005 Sprout Prize for the best book on international environmental policy and politics. Recent or current research efforts include an international comparison of twenty countries' domestic forest policy regulations (Earthscan); a comparative study on firms' responses to forest certification in the US forest sector (with Graeme Auld, Aseem Prakash and Erika Sasser); and an analysis (with Steven Bernstein) of the emergence of non-state market-driven global governance generally. He is author or coauthor of several articles that have appeared in the *American Journal of Political Science*, *Global Environmental Politics*, *Governance*, *Policy Sciences*, *Canadian Journal of Political Science*, *Regulation and Governance*, *Business and Politics*, *Forest Policy and Economics*, *Journal of Forestry*, *Canadian Public Administration*, *Canadian-American Public Policy* and *Forestry Chronicle*.

Isabelle Engeli is Assistant professor in public policy in the Graduate School of Public and International Affairs at the University of Ottawa. Her current research specializes on comparative policy agendas,

regulatory regimes of human biotechnology and gendering policy change. She recently codirected a comparative research project on morality issues across Western Europe (*Morality Politics in Western Europe: Parties, Agendas and Policy Choices,* 2012, with C. Green-Pedersen and L. Thorup Larsen). Her work has appeared in the *Journal of European Public Policy, West European Politics, Political Research Quarterly*, *Comparative European Politics*, *French Political Science Review* and *Swiss Political Science Review*. Isabelle is the corecipient of the 2012 APSA Best Comparative Policy Paper Award and the 2011 Carrie Chapman Catt Prize.

Fabrizio Gilardi is an associate professor at the Department of Political Science in the University of Zurich, Switzerland. His work on regulatory institutions and policy diffusion has appeared in the *American Journal of Political Science, British Journal of Political Science, Comparative Political Studies, Journal of European Public Policy* and *Journal of Theoretical Politics,* among others. His latest book, coauthored with Martino Maggetti and Claudio M. Radaelli, is *Designing Research in the Social Sciences* (2013).

Markus Haverland is Professor of Political Science at the Department of Public Administration, Erasmus University Rotterdam, and Fellow at the Montesquieu Institute in The Hague. His interests include the politics of EU policy making, the effect of the EU on member states, comparative public policy and the methodology and methods of causal case studies. He has published on these themes in journals such as the *Journal of Common Market Studies*, *Journal of European Social Policy, Journal of European Public Policy, Journal of Public Policy, Public Administration, Public Administration Review* and *West European Politics*. Together with Joachim Blatter, he has published *Designing Case Studies. Explanatory Approaches in Small-N Research* (2012).

Adrienne Héritier is Professor of Political Science in the Department of Political and Social Science and the Robert Schuman Center for Advanced Studies, European University Institute, Florence. Her research extends to theories of institutional change in the European Union, comparative public policy, European policy making, Europeanization, regulation and new modes of governance. Her books include *Changing Rules of Delegation* (2013), *Explaining Institutional Change in Europe* (2007), *Policy-Making and Diversity in Europe: Escape from Deadlock* (1999) and *Refining Regulatory Regimes: Utilities in Europe* (2006, coedited with David Coen). She has extensively published in various journals such as *West European Politics*, *Journal of European Public Policy*, *Business and Politics* and *Journal of Public Policy*.

Season Hoard is a postdoctoral research coordinator for the Division of Governmental Studies and Services and an adjunct instructor at Washington State University. Her research interests focus on comparative methodology, comparative public policy and women's political representation with an emphasis on Western postindustrial democracies.

Michael Howlett is Burnaby Mountain Chair in the Department of Political Science at Simon Fraser University and Yong Pung How Chair Professor at the Lee Kuan Yew School of Public Policy at the National University of Singapore, specializing in public policy analysis, political economy, and resource and environmental policy. He is the author of *Designing Public Policy* (2011), *European and North American Policy Change* (2009, coedited with G. Capano) and *Studying Public Policy* (2009, with M. Ramesh and A. Perl). His work has been published in the *American Journal of Political Science, Policy Sciences, Policy Studies Journal, Canadian Journal of Political Science* and *Journal of Public Policy.*

Amy G. Mazur is a professor in the Department of Political Science at Washington State University. Her research and teaching interests focus on comparative feminist policy issues with a particular emphasis on France. Her books include *Comparative State Feminism* (1995, coedited with Dorothy McBride), *Theorizing Feminist Policy* (2002) and *Politics, Gender and Concepts* (2008, coedited with Gary Goertz) and *The Politics of State Feminism: Innovation in Comparative Research* (2010, coedited with Dorothy McBride). She has published articles in *Comparative European Politics, Revue Française de Science Politique, Politics and Gender, Political Research Quarterly, Policy Studies Journal, West European Politics* and *European Journal of Political Research.*

Benoît Rihoux is Professor of Political Science at the Université catholique de Louvain (Belgium). His substantive research interests include political parties, new social movements, organizational studies, political change and policy processes. He co-ordinates the COMPASSS international network (http://www.compasss.org) on configurational comparative methods and QCA and has published extensively on their applications in diverse fields. He is also joint convener of international initiatives around methods such as the ECPR Methods School. He has published *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques* (2009, coedited with Charles Ragin).

Christine Rothmayr Allison is an associate professor in the Department of Political Science at the University of Montreal. Her fields of interest are comparative courts and politics and comparative public policy, where she studies policies in the field of biotechnology, biomedicine and higher education. Her recent research analyzes the agenda of the Canadian Supreme Court in comparative perspective and looks at how courts contribute to shape biotechnology policies in Europe and North America. Her books include *The Politics of Biotechnology: Policy Networks, Institutions and Internationalization* (2007, coedited with E. Montpetit and F. Varone). Her work has been published in the *European Journal of Political Research*, *Comparative Political Studies*, *Journal of European Public Policy* and *West European Politics*.

Jeroen van der Heijden is Assistant Professor of Environmental Governance at both the Australian National University (Regulatory Institutions Network, RegNet) and the University of Amsterdam (Amsterdam Law School). His general research interest is in regulation and governance. He is currently involved in a four-year research project (2012–2016) that seeks to understand the performance of innovative environmental governance arrangements that aim to improve urban sustainability and resilience. This project builds on close to 60 real-world examples of such arrangements in Australia, India, Malaysia, the Netherlands, the United States and Singapore. He has published widely on public sector reform, hybridization of governance, innovations in environmental governance and voluntarism in the fields of climate change and environmental policy. His work is generally comparative in nature. He has published in journals such as *Law & Policy*, *Regulation & Governance*, *International Political Science Review*, *Policy Science*, *Journal of Environmental Policy and Planning* and *Administration & Society.*

Joseph Wong is the Ralph and Roz Halbert Professor of Innovation at the Munk School of Global Affairs, University of Toronto, and Professor and Canada Research Chair in the Department of Political Science. He is also the Director of the Asian Institute at the Munk School and Senior Fellow at the Asia Pacific Foundation of Canada. Wong's research focuses on public policy and political economy in Asia. He is the author of *Betting on Biotech: Innovation and the Limits of Asia's Developmental State* (2011); *Healthy Democracies: Welfare Politics in Taiwan and South Korea* (2004) and coeditor (with Dilip Soman and Janice Stein) of *Innovating for the Global South: Towards an Inclusive Innovation Agenda* (2014) and (with Edward

Friedman) *Political Transitions in Dominant Party Systems: Learning to Lose* (2008). His articles have appeared in a broad range of journals including *Perspectives on Politics, Politics and Society, Governance, Comparative Political Studies, Pacific Affairs, Studies in Comparative International Development* and *Journal of East Asian Studies,* among others.

Dvora Yanow, Guest Professor in Wageningen University's Department of Social Sciences, Communication, Philosophy, and Technology Sub-department and Professor of Organizational Studies at Keele University, is a comparative policy analyst, policy/political and organizational ethnographer, and interpretive methodologist whose research and teaching are shaped by an overall interest in the generation and communication of knowing and meaning in organizational and policy settings. Current research engages comparative immigrant integration policies and citizen-making practices, state-created categories for race-ethnic identity, research regulation (ethics board) policies, practice theory and the life cycle, science museums and the idea of science, and built space/place analysis. Her most recent book, *Interpretive Research Design: Concepts and Processes* (2012), with Peregrine Schwartz-Shea, is the first volume in their coedited Routledge Series on Interpretive Methods. Her articles have appeared in the *Journal of International Relations and Development*, *Policy Sciences*, *Political Research Quarterly*, *Organization Studies*, *American Review of Public Administration* and *Journal of Management Studies*, among others.

1
Conceptual and Methodological Challenges in Comparative Public Policy

Isabelle Engeli and Christine Rothmayr Allison

The need for a more integrated methodological perspective in comparative public policy

Handbooks and manuals on public policy regularly open with a discussion on whether there is such a thing as a field of policy studies. Some point to the fact that the diversity of methodological and theoretical approaches, and the more 'craft-oriented' than purely science-oriented character of policy studies, make the building of grand theory a vain endeavour. Others, in order to affirm that we can meaningfully speak of a field of policy studies, have assembled and sometimes integrated concepts and theoretical frameworks in order to distinguish policy studies from other research programmes in the social sciences as a subdiscipline in its own right. This volume does not engage in this type of discussion, but takes for granted that there is a research tradition going back to the work of Lasswell (1951, 1970) that we can call policy studies. Lasswell conceived of the policy sciences as being problem focused, that is, interested in the substantive societal issues and problems facing governments which they need to address through analyzing the processes of policy formulation and choices, and by evaluating implementation and policy outcomes. He advocated a multidisciplinary, multimethod and theory-driven approach. In this approach, in order to contribute to problem solving, we need theories of the policy process in order to understand the mechanisms and factors that shape policy choices and policy outcomes. The ultimate goal of policy science for Lasswell was to contribute to the democratization of society.

This volume incorporates a specific interpretation of the Lasswellian approach to the policy sciences, and proposes a more precise focus on the central issue of methodological challenges in comparative policy

studies. The field of public policy analysis comprises broadly two different, related and equally valuable scholarly enterprises: to provide knowledge and policy expertise in and for policy making (*policy sciences, policy analysis*), and to develop general theories and frameworks of the policy process explaining and predicting policy-making processes (*policy studies*). The latter focuses on how problems are defined, agendas set, policies formulated, decided, implemented and evaluated (Parsons, 1995: p. XVI). Both research traditions embrace comparative research in order to ensure their findings contribute to better theories on policy making.

This book is part of the second scholarly endeavour, that of *policy studies*. Comparative policy studies address processes of policy making, of problem emergence and definition, of policy formulation, of policy implementation and also evaluation. Why governments choose different courses of action – or decide not to act at all – is the classical question of comparative policy studies, and constitutes a central aspect of the discussions in this volume (Heidenheimer et al., 1990: p. 3). Drawing on the seminal work of Heidenheimer et al. (1990), this volume places comparison at the heart of public policy research. Comparative analysis encourages moving beyond the particularities of each case and identifying patterns and regularity across cases, settings and time periods. Comparative designs force the researcher not to stop the analysis at particularistic explanations drawn from a single context, but to test whether the answers to research questions hold true for a larger number of cases and contexts.

For example, the policies designed to regulate the emerging field of human biotechnology provide a vivid illustration of the utility of comparison. Countries have adopted highly contrasting responses to the challenges posed by human biotechnology development and embryonic research. For instance, France launched the policy decision-making process on human biotechnology early in the 1980s. The proponents of early regulation emphasized the problems stemming from the legal void regarding human biotechnology at that time. On the contrary, one of France's neighbours, Belgium, formulated a completely different policy response to the same legal void and waited for more than 20 years to regulate the field. Understanding why, how and to what extent policy responses to human biotechnology development diverge requires a comparative research design (Varone et al., 2005; Montpetit et al., 2007; Engeli et al., 2012).

In the same vein, a comparative approach helps disentangle the morass of competing, and often conflicting, accounts of policy responses

to a common event, often conceived of in the literature as an exogenous 'shock to the system'. Contemporary examples of such events are the emergence of global terror networks or the global economic crisis of 2008. In both cases, governments scrambled to make policy in a compressed time frame, with limited antecedents or precedents, and in the face of great pressure from organized groups and the public for effective action. Much has been written from a descriptive perspective, often using narrative techniques, that has given insight into the forces at play within particular policy-making contexts (that is, the United States); however, the limitations of *sui generis* become clearly evident when cases are added and the analysis becomes comparative (Hajer, 2009; Culpepper, 2011). Comparative analysis widens the understanding of potential policy options – thus pushing the researcher to explain not only why certain policies were adopted, but also why others were not – and also permits the building of classifications and taxonomies of regulations, policies and instruments.

Why this book?

Since the seminal work of Heidenheimer et al. (1990), the field of comparative public policy has developed tremendously and the research community is ever expanding. With this expansion, new questions, new theories and new methodological challenges have emerged. In particular there has been a growing interest in conducting comparative policy studies for various reasons. The digital age together with higher standards in terms of government transparency and access to information, have increased the availability and accessibility of policy relevant information, data and statistics. Globalization and regional processes of economic integration, together with processes of decentralization or even federalization, have transformed national and international political institutions and pushed policy scholars to integrate concepts of internationalization, Europeanization and multilevel governance in their research and research designs in order to account, for example, for increasing interdependence. We can also observe a more critical approach to theories developed within the US institutional context that have been prominent in policy analysis and that scholars want to test and further develop in order to test their generalizability. Last, but not least, in various countries we can observe a 'comparative turn' in the study of national politics that has also contributed to strengthen comparative policy research. Independent of the reasons for the increased interest in comparative policy studies, policy scholars have turned to comparative designs in

order to be able to better identify patterns and regularities across cases, in order to go beyond the particularities of any single case, irrespective of the research tradition they identify with.

Despite a considerably greater emphasis on comparative policy research, policy scholars currently have no single text at their disposal that takes into account the various research traditions in policy studies and provides an overview over various methodological approaches from a public policy perspective. This volume is not intended to replace publications that discuss a specific type of research design, specific traditions of analysis or methods of analysis in detail. Similarly, it is not intended to provide the reader with fine grained knowledge on the methods of data collection and analysis. Rather it provides an overview of the basic design choices which researchers conducting comparative policy studies have to answer and encourages the more experienced researcher to reconsider their own research in the light of other methodological possibilities for designing comparative research. Hence, while emphasizing the importance of comparison in policy studies, from a methodological perspective this volume advocates a pluralist view. Each chapter is dedicated to a specify method that often relies on a particular type of data. Accordingly, each chapter also discusses the relevant data for the method presented. This way of organizing the volume allows for taking into account the diversity of data used in comparative policy analysis.

Addressing methodological challenges in comparative policy studies

Comparative policy studies face various methodological challenges. Researchers first have to conceptualize the object of comparison, public policy, and choose the cases that will be compared. The first part of the book discusses these two basic methodological challenges facing any research project. The second part familiarizes the researcher with the most widely used comparative designs, representing different traditions within policy studies, reaching from case studies and process tracing, to medium-N and large-N studies and addressing interpretivist analysis. The third part of the book takes up recent methodological developments in comparative policy, namely comparing beyond traditional European and North American cases, addressing multilevel governance, integrating gender into policy studies and using mixed-methods designs. The chapters in all three parts extensively cite examples of comparative policy studies in order to provide guidance for concretely designing research projects. Finally, it is worth pointing out that the stage of actually conceptualizing

public policy and choosing a research design is preceded by a more fundamental reflexion about what public policy studies are and ought to be. As the chapters assembled in this volume illustrate, an interpretivist approach to comparative public policy (*for example, Chapter 7*) would have a substantially different take on the conceptualization of public policy than would a gender-based approach (*for example, Chapter 10*) or a more classical postpositivist analysis (*for example, Chapter 6*).

Part I: Types of comparisons and their methodological challenges

The book starts out with defining the object of comparison, public policies, and proposes several conceptual angles to do so. As Howlett and Cashore point out in *Chapter 2*, the various ways in which policies have been defined poses a challenge for accumulating knowledge, building and testing theories across various comparative policy studies. As the authors demonstrate, one of the challenges of comparing public policies lies within the object of comparison itself, as '… public policies are complex entities made up of a number of constituent parts, since they exist as combinations of goals and means put together and implemented by a variety of authoritative policy actors operating within an environment of multiple interacting actors and organizations operating over both time and space' (Howlett and Cashore, this volume, p. 20).

Hence, researchers have to make conscious conceptual choices when designing a comparative policy study, as studying all phases of the policy cycle from agenda setting to policy evaluation – and this across time and space – constitutes a great challenge for collecting and analyzing relevant data. In fact, more often then not, theoretical discussions focus on specific stages of the policy cycle, that is, trying to explain why policies vary across countries in terms of goals and instruments, comparing policy agendas in order to understand shifts in attention, looking at processes of emulation or policy diffusion across countries or then looking at how policy discourse or policy frames structure policy output and implementation. While the stages approach can be helpful in order to conceptualize with more clarity what is being studied, it entails some drawbacks, notably by artificially separating stages that in reality overlap and are linked with each other (as the authors point out) through various feedback loops. Nevertheless, thinking along stages also points to how the actor constellation varies throughout the policy cycle, drawing our attention to the fact that the interaction of policy actors, which is often at the heart of understanding public policies, is embedded in a larger institutional context and institutions create different constraints and opportunities in comparative perspective.

According to Howlett and Cashore, the currently dominant approach addresses the complexity of public policies by focusing on policy *dynamics*, that is, how polices evolve over time. This approach has several implications for designing comparative policy research, as it postulates a sufficiently long period for studying change (ten years), to conceptually distinguish fundamental (paradigmatic) change from incremental changes, to attribute considerable weight to institutions in policy reproduction, and to point out mechanisms outside of policy subsystems explaining fundamental change.

The question of how to conceptualize public policies is the first question addressed in this collection that cuts across all the various research designs presented later on. The second issue that any comparative design has to address is the question of case selection, which is crucial in its consequences for the inferential process. Small- and intermediate-N comparative policy research designs, which are widely applied in comparative policy research, are particularly confronted with the challenges posed by case selection as they are dealing with a limited number of cases – due to limited availability in reality or to limited research resources – and therefore cannot rely on a randomized sample of cases. They are frequently subject to the 'Too few cases/Too many variables' (Goggin, 1986) problem that policy studies often face. *Chapter 3* by Jeroen van der Heijden discusses the two strategies of case selection that are widely used in comparative policy studies, most similar systems or most different ones. These two strategies, however, contain considerable limitations because of the over-determination of the dependent variable and because of the limited number of cases available. As the number of available 'real' cases is limited, comparative policy analysis may take advantage of combining two or more comparative lines of inquiry in order to strengthen the generalization capacity of their explanation: cross-country comparison, comparison across policy domains and comparison across time. This compound research design method, recently proposed by Levi-Faur (2004, 2006) aims at maximizing the explanatory capacity of qualitative comparative analysis. Starting with the assumption that there are four main levels of comparison in public policy analysis – national, sectoral, international and across time – Levi-Faur (2004, 2006) proposes to select cases, which must vary on at least two of the four levels and then to systematically test causal hypotheses iteratively across levels. The richness of qualitative analysis is preserved while, simultaneously, the capacity for generalization is strengthened.

Part II: Comparative designs

Part II presents different types of comparative designs, each constituting a varying tradition within policy studies. It offers an overview over the most widely used designs, starting with the most classical approach of single case and small-N studies that occupy a prominent place in comparative public policy (*Chapter 4*). It then moves to more recent methodological developments by discussing intermediate-N comparisons using configurational comparative methods (QCA) (*Chapter 5*). The greater availability of policy data and the larger number of comparative projects replicating research projects across a large number of countries (for example, the Policy Agendas Project and the Comparative Agendas Project) have spurred interest in large-N comparisons in recent years, constituting the third tradition of comparative designs discussed in the second part of the book (*Chapter 6*). The last chapter in *Part II* addresses interpretive policy studies and discusses its epistemological and methodological characteristics (*Chapter 7*).

Case studies have occupied a prominent place in comparative policy studies. More recently process tracing has increasingly gained prominence as a within-case method in public policy analysis. In *Chapter 4*, Joachim Blatter and Markus Haverland first review recent innovations proposed in case studies, such as critical cases, and then discuss the contribution of process tracing to causal explanation based on the investigation of causal mechanisms. Case studies and process-tracing analyses are commonly described as 'meaning centred', in the sense that they are focused on the thick understanding of a few cases. Instead of considering case studies as a 'by default' comparison when research resources for a large-N are lacking, the authors argue that process tracing offers a powerful tool to enhance theory building and, to a certain extent, theory testing in policy studies. Comparative case studies provide in-depth knowledge of each case which can be used, through process tracing, to simultaneously confront alternative explanations in order to determine which of them survive empirical testing in the most robust way. Starting the analysis with competing hypotheses offered in the literature, process tracing systematically examines the accuracy of rival explanations at each stage of the process being analyzed on the basis of triangulated sources of evidence.

From single and small-N comparisons, *Chapter 5* moves on to the issue of intermediate-N comparison with QCA. Due to limited diversity of cases in reality (that is, limited availability of public policies in a given

domain) or due to limited research resources (that is, research funding or language barrier), comparative policy analysis is frequently confronted by the principal challenge associated with intermediate-N comparison: How to benefit from the complexity of each case while enhancing the generalization across cases? The chapter by Isabelle Engeli, Benoît Rihoux and Christine Rothmayr Allison reviews the various QCA techniques and addresses the most recent innovations such as fuzzy sets and probabilistic assessment. QCA aims at going beyond methodological disputes over quantitative reasoning (attempting to generate generalized explanation) versus qualitative reasoning centred on either one or a small number of cases (aimed at understanding the specificities of a particular public policy). Specifically conceived for intermediate-sized comparison, QCA strengthens the explanatory potential of qualitative analysis by systematizing cross-case comparison without limiting the richness of within-case analysis. Based on algorithms derived from Boolean algebra, QCA simplifies the interaction of explanatory factors in order to identify configurations of causality valid across cases. Thus it allows for the development of parsimonious qualitative explanations with a strong capacity for middle-range generalization. QCA also sheds light on the phenomenon of multicausality, a perennial problem for policy studies; public policies often represent complex political phenomena, and they may not be fully captured by a unique, monotonic explanatory factor.

With the increased availability and accessibility of policy-relevant quantitative data, large-N studies are becoming more prominent in comparative policy analysis. *Chapter 6* presents quantitative comparisons, discusses the methodological implications of such approaches, and provides a critical discussion of their application to comparative policy studies. Quantitative comparison aims at explaining and predicting patterns of policy-making processes and policy outcomes across cases. Christian Breunig and John Ahlquist focus in particular on description, theory testing and prediction as the three main goals of quantitative comparative policy studies. For each goal, the authors present and discuss data and particular procedures from comparative welfare state research. Based on a large sample of cases, quantitative comparison attempts to broaden the generalization of research findings. While quantitative studies suffer less from case selection bias than small-N comparison, they nevertheless encounter a series of methodological challenges that Breunig and Ahlquist discuss in their chapter. Among other challenges, large-N comparison is particularly confronted with a high level of heterogeneity within the sample that may result in concept stretching. Large-N comparison is also frequently constrained,

due to limited research resources, to rely on aggregate data provided by external sources (such as national-level statistics), which may endanger the comparability of the data. In the particular context of developing countries, this aggregation problem may be further compounded by issues of reliability.

From a methodological point of view, research designs do not simply vary in terms of the number of cases and techniques of analysis and comparison they employ, but also in terms of their epistemological stance. Since the argumentative turn in policy analysis, critical policy studies and interpretative approaches are firmly anchored in the field and, equally so, in comparative policy analysis. Following the interpretive turn in the social sciences, a strand of research has emerged to promote a postpositivist perspective that places values, ideas and social meanings at the centre of public policy analysis. *Chapter 7* by Dvora Yanow introduces the reader to interpretive policy analysis. Going beyond a simple debate over analytical techniques, interpretative policy scholars call for a radical change in the aims of public policy studies in favour of an interpretative understanding of reality based on discursive and other modes of analysis. The chapter offers a critical discussion about the application of the interpretive turn to empirical analysis in the field of comparative policy studies, first reviewing the core assumptions of interpretive thought in policy studies more generally. Interpretivist methodology emphasizes the central role which social meanings and values play in the understanding of social realities and, because language is an important source of meanings, interpretive policy analysis engages discourses. The chapter discusses the implications of interpretivist methodology for research methods and presents a series of analytic approaches used, such as metaphors, categories and frames.

Part III: New challenges in comparative designs

Part III of the book takes up more recent developments in comparative policy studies and critically discusses their contributions, but also their limitations in terms of feasibility and pay-off. This section invites readers to strengthen their research designs by enlarging the regions and countries commonly compared (*Chapter 8*), more consciously addressing interdependence in multilevel governance (*Chapter 9*), considering gender biases in theory and methodology (*Chapter 10*) and profiting from mixed-method approaches when designing comparative policy research (*Chapter 11*).

Comparative policy studies traditionally focus on Europe, North America and Australia/New Zealand. More and more policy scholars are

also interested in analyzing how developing countries, countries with very different democratic traditions and economies, respond to the same policy challenges – or how policies diffuse across the globe and not just in Europe and North America. However, there is a lack of systematic discussion about the benefits and methodological challenges and potential of this type of comparison in the current literature. Joe Wong in *Chapter 8* takes up this challenge and discusses why policy scholars should go beyond classical comparisons; he then analyzes the challenges of such comparisons in terms of the choice of cases (small-N and/ or larger quantitative studies), practical questions of access to data and data analysis, and interpretation of results. He argues that there are no great conceptual or methodological adjustments required to compare unfamiliar cases, as the challenge in comparing beyond the Anglo-European world is one of unfamiliarity rather than incomparability. The chapter first recalls the classical debate in comparative analysis on conceptual rigour and the pitfalls of conceptual stretching, and discusses issues of data collection and fieldwork. The argument focuses on the ontological gap between the analyst's *expectations* and *empirical reality* and invites the policy analyst to consider (i) the implications of contextualized rationality; (ii) the importance of context-specific heuristic cues; and (iii) those prevailing theoretical biases which may lead us to infer incorrect conclusions. When comparing beyond familiar cases, researchers need to be aware of their own ontological predispositions and biases.

In an age of globalization and regional economic integration, researchers are not only faced with the issue of broadening the traditional spectrum of countries and regions compared, but multilevel governance also poses particular challenges to comparative policy studies. 'Cases' and levels are not independent from each other, because relevant actors are not limited to state-actors, and because multilevel governance is characterized through overlapping networks. Fabrizio Gilardi, in *Chapter 9* highlights the interest of interdependence as a methodological challenge to comparative policy studies. He discusses in detail the quantitative and qualitative strategies that researchers can deploy to account for interdependence when analyzing policy diffusion, policy transfer and policy convergence. Diffusion and transfer share the basic idea that policy choices in one unit are shaped by policy choices in other units, while policy convergence, the increased similarity of policies over time, is not necessarily related to interdependence. Spatial regression and dyadic analysis are quantitative methods that account for interdependence when studying processes of diffusion. Researchers interested in

uncovering evidence on the mechanisms driving processes of diffusion and transfer rely on qualitative methods and in particular within-case analysis and process tracing. The chapter advocates the combination of quantitative and qualitative methods in order to further develop the study of policy interdependence.

A quite different type of methodological issue is addressed in *Chapter 10* in the form of gendering comparative policy studies. Amy G. Mazur and Season Hoard tackle the growing importance of the gender dimension in comparative public policy analysis. Despite the fact that it is often hidden behind rationalization, the choice of policy goals and policy instruments is based on social representations such as gender roles and the definition of women's and men's identity. This chapter introduces the reader to the contribution and the methodological implications of gendering policy analysis. The two authors demonstrate that Feminist Comparative Policy (FCP) fills crucial empirical gaps in our knowledge about the state and policy in postindustrial democracies and, in doing so, places issues of democratic governance and performance to the fore. It also argues that non-feminist policy scholars need to better integrate FCP study designs, methods and findings in order to make comparative policy studies more systematic and, ultimately, to do better science. After introducing the core features and most recent methodological contributions and issues of FCP, the chapter addresses its four major streams; gendering welfare states, feminist policy formation, women's movements and policy and state feminism. They then discuss three new and recent trends, namely representation, intersectionality and feminist institutionalism. These research streams are presented in terms of how they gender comparative policy analysis and the added value they bring to the conduct of policy research in comparative perspective.

In recent decades, efforts to bridge the quantitative-qualitative divide in political science and the social sciences more generally have successfully drawn greater attention to mixed-method designs. Much research in comparative policy analysis relies on a single methodology, using either qualitative or quantitative tools. Recently a growing number of studies demonstrate the benefits of combining methods by including multiple pathways to knowledge and assessment in order to tackle better the complexity of public policy. Adrienne Héritier and Sophie Biesenbender (*Chapter 11*) first present the aims of mixed-methods thinking and then map out the various forms of mixed-methods designs. One may simultaneously combine different methods and techniques across the stages of the research and triangulate data to ensure validation of the results. Alternatively, one may first conduct a

quantitative phase of the study in order to get a systematic overview of the policy under study that is then complemented by a qualitative stage that provides deeper knowledge of some particular aspect(s) of it. The chapter then assesses the advantages and the weaknesses of mixed-methods designs with the support of concrete empirical applications. Mixed-methods designs lay out several advantages, such as looking at a broader set of the aspects of the research question as well as corroborating evidence; at the same time they are time consuming and require solid grounding in both quantitative and qualitative methods.

To summarize, the chapters assembled in this volume address the major methodological challenges that today's scholars of comparative policy face. Inevitably, they will encounter others in their research, but this volume highlights the most pressing of them that need to be addressed in the process of research design. Individual chapters provide concise instructions for the identification of and remedies to specific obstacles to good research design in comparative policy studies; in addition, concrete examples are provided, from extant and published works, that show how to operationalize these recommendations in a concrete manner. Taken together, the chapters provide answers to the major methodological challenges, and give a guide to constructing rigorous and analytical work in comparative policy studies. Readers will not only engage in the more abstract and conceptual debates over research design in the study of comparative public policy, but they will also be guided through the definition of research design, data gathering and comparative analysis. Ultimately the volume aims at contributing to the improvement of the design and execution of research programmes in comparative policy studies; in doing so, it seeks to augment theory building and theory testing in the field.

Bibliography

Culpepper, Pepper D. (2011) *Quiet Politics and Business Power: Corporate Control in Europe and Japan* (Cambridge University Press).

Engeli, Isabelle, Christopher Green-Pedersen and Lars Thurop Larsen (2012) *Morality Politics in Western Europe. Parties, Agenda and Policy Choices* (London: Palgrave Macmillan).

Goggin, Malcolm L. (1986) 'The "Too Few Cases/Too Many Variables" Problem in Implementation Research', *Western Political Quarterly* 38, 328–47.

Hajer, Maarten A. (2009) *Authoritative Governance: Policy-making in the Age of Mediatization* (Oxford University Press).

Heidenheimer, Arnold, Heclo Hugh and Carolyne T. Addams (1990) *Comparative Public Policy: The Politics of Social Choice in America, Europe and Japan*, 3rd edn (New York: St Martin's Press).

Lasswell, Harold D. (1951) 'The Policy Orientation', in Daniel Lerner and Harold D. Lasswell (eds) *The Policy Sciences: Recent Developments in Scope and Method* (Stanford University Press), 3–15.

Lasswell, Harold D. (1970) 'The Emerging Conception of the Policy Sciences', *Policy Sciences*, 1 (1), 3–14.

Levi-Faur, David (2004) 'Comparative Research Designs in the Study of Regulation: How to Increase the Number of Cases Without Compromising the Strengths of Case-Oriented Analysis', in Jacint Jordana and David Levi-Faur (eds) *The Politics of Regulation: Institutions and Regulatory Reforms for the Age of Governance* (Cheltenham: Edward Elgar), 177–99.

Levi-Faur, David (2006) 'Varieties of Regulatory Capitalism: Getting the Most Out of the Comparative Method', *Governance* 19 (3), 367–82.

Montpetit, Eric, Christine Rothmayr and Frederic Varone (2007) *The Politics of Biotechnology in North America and Europe: Policy Networks, Institutions and Internationalization* (Lanham: Lexington).

Parsons, Wayne (1995) *Public Policy: An Introduction to the Theory and Practice of Policy Analysis* (Aldershot: Edward Elgar).

Varone, Frederic, Christine Rothmayr and Eric Montpetit (2006) 'Regulating Biomedicine in Europe and North America: A Qualitative Comparative Analysis', *European Journal of Political Research* 45 (2), 317–43.

Part I
Types of Comparisons and Their Methodological Challenges

2
Conceptualizing Public Policy
Michael Howlett and *Ben Cashore*

Defining public policy and the need for comparative research

Policy-making involves both a technical and political process of articulating and matching actors' goals and means. Policies are thus actions which contain goal(s) and the means to achieve them, however well or poorly identified, justified, articulated and formulated. Probably the best-known, simple and short definition of public policy has been offered by Thomas Dye, 'anything a government chooses to do or not to do' (Dye, 1972: 2).While many organizations and actors create policies to which their members must adhere, we focus on 'public' policies made by governments that affect and influence every member of a nation-state or a subnational jurisdiction. The 'actions' we are concerned with in this case are government decisions to act, or not to act, to change or maintain some aspect of the status quo (Birkland, 2001: ch. 1).

Dye's definition specifies clearly that the primary agent of public policy-making is a government, rather than private business decisions, decisions by charitable organizations, interest groups, or other social groups. Although these might be important actors with some role in governmental policy-making processes, governments enjoy a special status in public policy-making due to their unique ability to make *authoritative* decisions on behalf of citizens; that is, decisions backed up by the potential for sanctions for transgressors in the event of non-compliance. Of course, the activities of non-governmental actors may, and very often do, influence governments' policy decisions, and governments will sometimes leave the implementation or some other aspect of policy-making to non-governmental organizations (NGOs). But governments maintain control over whether, when and how other actors may be involved.

Thus, for example, how the medical profession interprets the causes of lung cancer and the solutions it proposes for prevention and cure may have a bearing on what a government eventually does about such a problem in terms of health-care policy. However, the profession's proposed solution to the problem is not itself a public policy; only measures that a government actually adopts or endorses – such as a ban on the sale or use of tobacco – actually constitute *public* policy.

Second, as Dye notes, public policy is, at its simplest, a choice made by government to undertake some course of action. Dye highlights the fact that public policy-making involves a fundamental choice on the part of governments to do something or to do nothing with respect to a problem and that this decision is made by government officials, be they elected or appointed politicians, judges or administrators. Thus a 'negative' or 'non-decision', or a government's decision to do nothing and simply maintain the current course of action or status quo (Crenson, 1971; Smith, 1979), is just as much a policy decision as a choice to alter it. Such 'negative' decisions, however, like more 'positive' ones, must be *deliberate*, such as when a government decides not to increase taxes or declines to make additional funds available for arts, health care or some other policy area. The fact we have the freedom to paint the interiors of our homes in colours of our choice, for example, does not mean that this is a public policy, because the government never deliberately decided not to restrict our options in this area.

Third, and closely related to this, Dye's definition also highlights the fact that a public policy is a *conscious* choice of a government. That is, government actions and decisions very often involve *unintended consequences*, such as when an effort to regulate tobacco consumption or some other vice results in the activity 'going underground' and operating illegally as a 'black' market. Unless this subsequent activity or consequence was specifically anticipated and intended by government (such as occurs when governments increase gasoline taxes to discourage automobile use and thus indirectly promote the use of public transit), the unintended consequences are not public policy but merely its unexpected by-product, which may sometimes be beneficial and sometimes not.

While Dye's three points are central to understanding that public policies emerge as the result of governmental decision-making processes, it is less clear how such decisions are arrived at or implemented. Other more complex definitions, such as that put forward by Jenkins (1978), offer a more precise conceptualization of public policy which also addresses these aspects of the subject. Here policy is defined as 'a set

of interrelated decisions taken by a political actor or group of actors concerning the selection of goals and the means of achieving them within a specified situation where those decisions should, in principle, be within the power of those actors to achieve'.

This definition is helpful in specifying the *content* of a policy decision, as being composed of the 'selection of goals and means' as noted above, although this says nothing about the actual selection process. Viewing policy as the pursuit of conscious goals nevertheless raises the significance for policy-making of the ideas and knowledge held by policy actors about policy goals and the tools or techniques used to achieve them, since these shape their understanding of policy problems and the 'appropriateness' of potential solutions to them.

Moreover, Jenkins' definition is also helpful in clarifying some of the implicit components of Dye's definition, which could be construed as limiting policy-making to a single choice: opportunity and result. Jenkins' definition instead presents policy-making as a *dynamic* process and explicitly acknowledges that it is usually the result of 'a set of inter-related decisions' that cumulatively contribute to an outcome rather than a single decision. Thus, a health policy, for example, consists of a series of decisions – about building health facilities, certifying personnel and treatment, and financing health-care provision, among many other related items – often taken over a lengthy period of time (Tuohy, 1999). This also highlights the complexity of the nature of the actors involved in policy-making as these various interrelated decisions are often made by different individuals and agencies within government, such as a Department of Health as well as Ministries of Finance or Social Welfare, and by various divisions and sections within them, and also can involve a much larger set of non-state actors than often assumed, resulting in a much more complex policy-making process than a quick reading of Dye's definition might at first suggest.

Jenkins also improves upon Dye's definition by adding the idea that a government's *capacity* to implement its decisions is a significant component of public policy and a major consideration affecting the types of action that government will consider. This recognizes the limitations on a government's ability to act that can constrain the range of options considered in particular decision-making circumstances and contribute to the success or lack of success of their efforts. A government's choice of a policy may be limited, for instance, by the presence or lack of financial, personnel or informational resources, by international treaty obligations or by domestic resistance to certain options, all affecting what kinds of actions are considered 'effective', 'feasible' and 'appropriate' in dealing

with an issue (Huitt, 1968; Meltsner, 1972; Majone, 1975; May, 2005). Thus, for example, we will not understand health policy in many countries without realizing the powerful opposition that the medical profession can mount against any government's effort to control health-care costs which might reduce the profession's income (Alford, 1972) or without taking into account the kinds of financial resources governments have at their disposal in providing health care and/or paying doctors, nurses and other health-care providers for their services. Similarly, understanding domestic government actions increasingly requires detailed awareness of the limits placed upon them, and the opportunities provided to them, by international agreements, treaties, and conventions (Doern et al., 1996; Milner and Keohane, 1996).

Specifying the constituent elements of public policy

As this definitional discussion suggests, public policies are complex entities made up of a number of constituent parts, since they exist as combinations of goals and means put together and implemented by a variety of authoritative policy actors interacting within an environment of multiple actors and organizations over both time and space.

   Understanding how policy-making processes work, and with what results, can be undertaken at the domestic or international levels; and analyses in the policy sciences are often undertaken in this fashion. However, increasingly, comparative studies of processes and outcomes across different sectors – such as health policy or environmental or industrial policy – as well as across different nations are called for and have been undertaken in the effort to generate better descriptions of processes and identify the factors which influence them. An example of a path-breaking effort to specify exactly what these constituent parts entail and how they can be put together in an empirically and conceptually rich fashion which draws on comparative analysis can be found in Peter Halls' work (1989, 1993) comparing the development of economic policies in western countries. This work distinguished between three basic elements or components of public polices: more or less abstract or general policy goals, the more concrete policy instruments used to implement them and the even more specific operational settings or calibrations used when these instruments are deployed.

   Although Hall himself suggested that only these three different components existed, the distinction he drew within the three different 'levels' of specificity of goals and means – from 'abstract' to 'concrete' and 'specific' – means it is possible to discern as many as six elements that go into making a public policy (see Table 2.1) (Liefferink, 2006).[1]

Table 2.1 A modified taxonomy of policy components following Hall (1989, 1993)

		Policy content	
	High-level abstraction	*Programme-level operationalization*	*Specific on-the-ground measures*
Policy ends or aims	**Goals** **What general types of ideas govern policy development?** (e.g. environmental protection, economic development)	**Objectives** **What does policy formally aim to address?** (e.g. saving wilderness or species habitat, increasing harvesting levels to create processing jobs)	**Settings** **What are the specific on-the-ground requirements of policy** (e.g. considerations about the optimal size of designated stream-bed riparian zones, or sustainable levels of harvesting)
Policy focus			
Policy means or tools	**Instrument logic** **What general norms guide implementation preferences?** (e.g. preferences for the use of coercive instruments, or moral suasion)	**Mechanisms** **What specific types of instruments are utilized?** (e.g. the use of different tools such as tax incentives, or public enterprises)	**Calibrations** **What are the specific ways in which the instrument is used?** (e.g. designations of higher levels of subsidies, the use of mandatory vs. voluntary regulatory guidelines or standards)

Note: Cells contain examples of each measure.
Source: Modified from Cashore and Howlett (2007).

Using this insight, public policies and public policy-making can be seen to revolve around the process of articulating and matching policy goals with preferred policy means at all three levels: abstract (general or conceptual), programme (concrete) and on-the-ground (settings). Thus, for example, in an area such as criminal justice policy, policy-making involves both consideration of general abstract policy goals (like reducing crime) and conceptual means (better policing), as well as programme-level objectives (such as reducing violent or gun-related crime) and mechanisms (by increasing the number of local police stations in high crime areas), and also the settings and calibrations of policy tools (such as reducing violent crime by 50 per cent over five years and, in doing so, doubling the number and frequency of police patrols in affected areas).

Describing how policy is made: The policy cycle model

Comparative studies of policy-making often focus on one or two of these elements of policy-making, but accurate depictions of policy processes and outcomes require investigation and analysis of all elements across these three levels and of the manner in which both goals and means are articulated (Bannink and Hoogenboom, 2007; Kuhner, 2007).

Works in this vein have developed several key insights into policy-making processes and behaviour which inform contemporary comparative policy studies. One is the finding that policy processes generally unfold as a set of interrelated stages through which deliberations concerning some issue or problem flow in a more or less sequential fashion from being an 'input' to government deliberations to being an 'output' or subject of government action. The sequence of stages through which this decision-making activity operates is often referred to as the '*policy cycle*' (Jann and Wegrich, 2007), an idea first broached in the work of Harold Lasswell (1956), one of the central pioneers and promoters of what he termed 'the policy science' (Farr et al., 2006). Lasswell (1971) divided the policy process into seven stages, which, in his view, described not only how public policies were actually made but also how they should be made.

In Lasswell's construct, the policy process began with intelligence-gathering, that is, the collection, processing and dissemination of information for those who participate in decision-making. It then would move to the promotion of particular options for addressing the problem by those involved in making the decision. In the third stage, the decision-makers would prescribe a course of action. In the fourth stage, the prescribed course of action would be invoked alongside a set of sanctions to penalize those who fail to comply with these prescriptions. The policy would then

be applied by the courts and the bureaucracy and run its course until it was terminated or cancelled. Finally, the results of the policy would be appraised or evaluated against the original aims and goals.

This way of thinking about policy-making as a staged process of problem-solving was highly influential in the development of the policy sciences (deLeon, 1999) and formed the basis for further work on the subject (Lyden et al., 1968; Brewer, 1974; Simmons et al., 1974; Anderson,1975; Jones, 1984). Through comparative studies of policy-making processes in many sectors and jurisdictions, a simpler, more parsimonious version of the policy cycle emerged which more clearly linked the stages of public policy-making with the logic of applied decision-making raised by Dye and others (Brewer and deLeon 1983; Hupe and Hill, 2006). The five stages in applied problem-solving and the corresponding stages in the policy process are depicted in Table 2.2.

In this model, *agenda setting* refers to the process by which problems come to the attention of governments; *policy formulation* refers to how policy options are formulated within government; *decision-making* is the process by which governments adopt a particular course of action or non-action; *policy implementation* relates to how governments put policies into effect; and *policy evaluation* refers to the processes by which the results of policies are monitored by both state and societal actors, the outcome of which may be a re-conceptualization of policy problems and solutions.

Much comparative policy study has focussed on detailing the operation of particular stages of the cycle; for example, examining through comparative case study techniques the nature of agenda setting dynamics in the United States and Europe, or comparing the roles played by specific actors – like the media – in each. Other studies examined the manner in which different stages interacted – such as feedback-like processes involving policy evaluation and agenda setting when a negative evaluation of a policy leads to its revision in subsequent rounds of policy-making (Pierson, 1993). These studies have shed a great deal of light on the nature of policy-making processes and highlighted the use of comparative

Table 2.2 Five stages of the policy cycle and their relationship to applied problem-solving

Applied problem-solving	Stages in policy cycle
1. Problem recognition	1. Agenda setting
2. Proposal of solution	2. Policy formulation
3. Choice of solution	3. Decision-making
4. Putting solution into effect	4. Policy implementation
5. Monitoring results	5. Policy evaluation

research as a technique to better understand the kinds of actors and dynamics at work in different countries and sectors as policy-making processes unfold.

Who makes policy: Policy actors

The cycle or process model of public policy-making is useful not only because of the way it separates a series of tasks conducted in the process of public policy-making into distinct phenomena which can be studied separately. It also helps clarify the different, but interactive, roles played at each point in the process by specific kinds of policy actors, the institutions in which they operate and the importance of the ideas they hold about both policy content and processes in determining what kinds of policy goals and means are considered and implemented (Sobeck, 2003).

Comparative study has helped show, for example, how at the agenda setting stage virtually any (and all) policy actors can be involved in decrying problems and demanding government action. These policy actors – whether all, many or few – can be termed the *policy universe*. At the next stage, formulation, however, research has shown how only a subset of the policy universe – the *policy subsystem* – is typically involved in discussing options to deal with problems recognized as requiring some government action. This subsystem is composed only of those actors with sufficient knowledge of a problem area, or a resource at stake, to allow them to participate in the process of developing alternative and feasible courses of action to address the issues raised at the agenda setting stage. When a decision is being taken to adopt one or more, or none, of these options and implement it, the number of actors is reduced even further, to only the subset of the policy subsystem composed of *authoritative government decision-makers*, whether these are elected or appointed officials, judges or bureaucrats. Once implementation begins, however, the number of actors increases once again to the relevant *subsystem* and then, finally, with the evaluation of the results of that implementation, expands once again to encompass the entire *policy universe*. This public policy 'hourglass' configuration of actors is set out in Table 2.3 below.

Table 2.3   The policy cycle–actor hourglass

Stages in policy cycle	Key actors involved
1. Agenda setting	1. Policy universe
2. Policy formulation	2. Policy subsystem
3. Decision-making	3. Authoritative government decision-makers
4. Policy implementation	4. Policy subsystem
5. Policy evaluation	5. Policy universe

Understanding how these different actors interact in policy processes in order to produce specific kinds of policy outcomes has been a major subject of contemporary comparative policy research. Different patterns of policy outcomes have been linked to the patterns of behaviour of different policy actors at each stage of the policy process and these, in turn, have been linked to factors such as the kinds of institutional structures found in different countries and sectors that condition how policy initiatives emerge and how policy advice is generated and deployed (Aberbach and Rockman, 1989; Bennett and McPhail, 1992; Bevir and Rhodes, 2001; Peled, 2002; Bevir et al., 2003; Howlett and Lindquist, 2004).

Recent comparative studies of policy formulation processes in New Zealand, Israel, Canada and Australia, for example, have developed the idea that government decision-makers sit at the centre of a complex web of policy advisors who are key players in affecting how demands made at the agenda setting stage of the policy process are articulated into specific policy options or alternatives for decision-makers to consider (Maley, 2000; Peled, 2002; Dobuzinskis et al., 2007; Eichbaum and Shaw, 2007). These include both 'traditional' political advisors in government as well as non-governmental actors in NGOs, think tanks and other similar organizations, as well as less formal or professional forms of advice from colleagues, relatives and members of the public.

Given its reliance on existing institutional configurations, the exact configuration of an advisory system varies not only temporally, but also spatially, by jurisdiction, especially by nation-state and, somewhat less so, by policy sector. That is, the personal and professional components of the policy advice supply system, along with their internal and external sourcing, are combined in different ratios in different situations (Prince, 1983; Wollmann, 1989; Hawke, 1993; Halligan, 1995; Rochet, 2004). Discerning the underlying patterns of policy analysis, their influence and effectiveness in different analytical contexts involves understanding how a policy advice system is structured and operated in different countries, jurisdictions and sectors of policy activity (Lindquist, 1998; Mayer et al., 2004; Verschuere, 2009) and is thus a good example of a subject well suited to comparative public policy analytical techniques.

How policies change: The punctuated equilibrium model of policy dynamics

These insights into the nature of public policy, policy processes and policy actors have been brought together in recent years by students of comparative public policy *dynamics*; that is, in order to explore and explain the manner in which policies change. In this regard, the

contemporary study of policy dynamics owes a broad debt to two works which appeared 30 years apart: Peter Hall's above-mentioned study of policy paradigms (Hall, 1989) and Charles Lindblom's earlier work on incrementalism (Lindblom, 1959).

Both authors worked in a synthetic fashion, utilizing the insights of other scholars into aspects of politico-administrative behaviour – in Lindblom's case Herbert Simon's insights into the nature of organizational behaviour (Simon, 1957), and in Hall's case Thomas Kuhn's ideas about the history of scientific advance (Kuhn, 1962) – to propose and refine the notion that general patterns of policy development can be identified and understood which transcend specific subject areas and content.

Hall's work served to break the long-term orthodoxy in studies of policy dynamics dominated by Lindblom-inspired incrementalism, that is, the position that due to constraints on the nature of human cognition which lead to most decision-making effectively being political bargaining, a single type of policy dynamics – marginal or small increments from the status quo – would characterize almost all instances of public policy-making (Hayes, 1992; Howlett and Ramesh, 2003). Although Hall agreed with this basic characterization and finding, he also identified a second pattern of change. This was the broad 'paradigm' shift which could be observed in many policy areas in many countries and sectors over long periods of time in which policy goals changed in a non-linear fashion; such as when ideas about 'wellness' in the health sector took over from long-held notions about 'illness'.

Comparative scholars studying public policy dynamics have been involved in a 20-year effort to better refine the two patterns and assess their interrelationship. As this work has developed, a new 'post-incremental' orthodoxy has emerged as policy scholars have generally accepted the idea – borrowed from paleo-biology (Eldredge and Gould, 1972) and first put forward in the context of policy dynamics by Baumgartner and Jones (1991) – that periods of marginal adaptation and revolutionary transformation are typically linked in an overall 'punctuated equilibrium' pattern of policy change. That is, they are linked in a pattern in which periods of relatively long-term stability in policy goals succeed each other through shorter unstable transitionary phases.

The principle elements of the new orthodoxy in comparative studies of policy dynamics are that:

- Any analysis of policy development must be historical in nature and cover periods of years or even decades or more (Sabatier and Jenkins-Smith, 1993).[2]

- Political institutions and their embedded policy subsystems act as the primary mechanisms of policy reproduction (Clemens and Cook, 1999; Botcheva and Martin, 2001; Howlett and Ramesh, 2003).
- 'Paradigmatic' change, a process in which there is a fundamental realignment of most aspects of policy development, is generally understood to occur rarely, and in the absence of such processes policy changes are expected to follow 'incremental' patterns (Genschel, 1997; Deeg, 2001); and,
- Paradigmatic transformations or 'punctuations' themselves usually occur due to the effects of 'external perturbations' that cause widespread disruptions in existing policy ideas, beliefs, actors, institutions and practices rather than due to endogenous causes, although this is also possible through processes such as policy learning (Pierson, 1993, 2000; Sabatier and Jenkins-Smith, 1993; Mahoney, 2000; Smith, 2000; Thelen, 2003, 2004; Hacker, 2004;Howlett and Rayner, 2006; Kay, 2006).

More comparative research is required in order to further refine these ideas (John and Margetts, 2003). For example, some recent longitudinal studies have failed to find evidence of the exogenously- or externally-driven change processes typically associated with most contemporary theorizing about paradigmatic changes (Coleman et al., 1996; Cashore and Howlett, 2007). In some cases, researchers have found dramatic policy change to involve a more complex pattern of linkages and change among the levels or orders of policy than originally contemplated (Mortensen, 2005). Uncovering these 'hidden' and more complex patterns of policy development challenges the way most policy studies measure and classify overall policy dynamics as either 'paradigmatic' or 'incremental' and provide another fertile field for contemporary comparative policy research (Howlett and Ramesh, 2002; Lindner, 2003; Lindner and Rittberger, 2003).[3]

Conclusion

The complexity of public policy is considerable, posing analytical difficulties for which students of comparative public policy-making have developed useful models or analytical frameworks such as those set out above (Yanow, 1992; Danziger, 1995; Phillips, 1996). The results of such comparative efforts have been many and fruitful. These comparative studies of policy elements, processes, actors and dynamics have shown public policy to be a complex phenomenon consisting of numerous decisions made by many individuals and organizations inside government at different points in policy processes, influenced by others operating within, and

outside of, the state and resulting, generally, in long periods of stability of outcomes or incremental changes, punctuated by infrequent bursts of paradigmatic change. The decisions policy-makers make have been shown to be shaped both by the structures within which these actors operate and the ideas they hold – forces that have also affected earlier policies in previous iterations of policy-making processes and have set policies onto specific trajectories, sometimes over long periods of time.

Thanks to these studies, the discipline now has a much stronger understanding about factors such as legislative 'attention spans', 'policy windows' and how alterations in subsystem beliefs and membership can result in certain issues coming to the fore on policy agendas, precipitating change by shaping what subsystem members deem to be appropriate types and modes of policy-making (Hall, 1989; Baumgartner and Jones, 1993, 2002; Kingdon, 1995; Leach et al., 2002). It also has a much better understanding of the role played by macro-, meso- and micro-institutions which formalize issue discourses and routinize political and administrative affairs, shaping the mobilization of actors and restraining change in policy agendas and processes (Weaver and Rockman, 1993; Thelen, 2003, 2004; Deeg, 2005).

However as pointed out above, this research agenda is by no means finished. A large number of significant questions about policy-making and policy behaviour continue to exist and define the subjects for further comparative research. In addition to the kinds of questions about patterns of policy change mentioned above, it is also the case that most studies have focussed on a small number of North American and European cases and many countries and sectors outside this region have received little attention. Whether and to what extent they share the same characteristics as oft-studied cases is uncertain. Similarly many studies have focussed on only a few well-documented, high profile policy sectors such as finance, trade or health. Many others still require detailed examination in order to see how well concepts developed in other sectors travel to them. And while spatial comparisons are reasonably plentiful, temporal ones lag far behind.

Thus, while comparative research has greatly enriched our understanding of policy-making, more and better comparative studies of policy-making over both time and space are still required to better inform both contemporary policy theorizing and practice.

Notes

1   For similar models based on a similar critique of Hall, see Daugbjerg (1997) and Smith (2000). These six categories are inspired from much of the work on applied policy analysis that teach students to break policy down into their 'goals', 'operationalized' objectives, and specific criteria and who likewise take

pains to distinguish policy instruments from 'on-the-ground' policy require-ments (Weimer and Vining, 1999).

2   This observation is explicitly raised by Baumgartner and Jones on punctuated equilibrium and in Paul Sabatier and Hank Jenkins-Smith's work on 'advocacy coalitions', as well as being implicit in the broad field of historical institution-alism (Sabatier, 1988; Sabatier and Jenkins-Smith, 1993; Mahoney, 2000; Lindner and Rittberger, 2003).

3   Both incremental and paradigmatic changes generally remain underspecified entities (Berry, 1990; Bailey and O'Connor, 1975; Kuhn, 1974; Capano, 2003).

Bibliography

Aberbach, Joel D. and Bert A. Rockman (1989) 'On the Rise, Transformation, and Decline of Analysis in the US Government', *Governance*, 2 (3), 293–314.

Alford, Robert R. (1972) 'The Political Economy of Health Care: Dynamics Without Change', *Politics and Society* 2 (2), 127–64.

Anderson, James E. (1975) *Public Policymaking* (New York: Praeger).

Bailey, John J. and Robert J. O'Connor (1975) 'Operationalizing Incrementalism: Measuring the Muddles', *Public Administration Review*, 35 (January–February), 60–6.

Bannink, Duco and Marcel Hoogenboom (2007) 'Hidden Change: Disaggregation of Welfare Regimes for Greater Insight into Welfare State Change', *Journal of European Social Policy*, 17 (1), 19–32.

Baumgartner, Frank R. and Bryan D. Jones (1991) 'Agenda Dynamics and Policy Subsystems', *Journal of Politics*, 53 (4), 1044–74.

Baumgartner, Frank R. and Bryan D. Jones (1993) *Agendas and Instability in American Politics* (University of Chicago Press).

Baumgartner, Frank R. and Bryan D. Jones (2002) *Policy Dynamics* (University of Chicago Press).

Bennett, Scott and Margaret McPhail (1992) 'Policy Process Perceptions of Senior Canadian Federal Civil Servants: A View of the State and Its Environment', *Canadian Public Administration*, 35 (3), 299–316.

Berry, William T. (1990) 'The Confusing Case of Budgetary Incrementalism: Too Many Meanings for a Single Concept', *Journal of Politics*, 52 (1), 167–96.

Bevir, Mark and Rod A. W. Rhodes (2001) 'Decentering Tradition: Interpreting British Government', *Administration & Society*, 33 (2), 107–32.

Bevir, Mark, Rod A. W. Rhodes and Patrick Weller (2003) 'Traditions of Governance: Interpreting the Changing Role of the Public Sector', *Public Administration*, 81(1), 1–17.

Birkland, Thomas A. (2001) *An Introduction to the Policy Process; Theories, Concepts, and Models of Public Policy Making* (Armonk: M.E. Sharpe).

Botcheva, Liliana and Lisa L. Martin (2001) 'Institutional Effects on State Behaviour: Convergence and Divergence', *International Studies Quarterly*, 45 (1), 1–26.

Brewer, Gene D. (1974) 'The Policy Sciences Emerge: To Nurture and Structure a Discipline', *Policy Sciences*, 5 (3), 239–44.

Brewer, G. and Peter deLeon (1983) *The Foundations of Policy Analysis* (Homewood: Dorsey).

Capano, Giliberto (2003) 'Administrative Traditions and Policy Change: When Policy Paradigms Matter, The Case of Italian Administrative Reform During the 1990's', *Public Administration*, 81 (4), 781–801.

Cashore, Benjamin and Michael Howlett (2007) 'Punctuating Which Equilibrium? Understanding Thermostatic Policy Dynamics in Pacific Northwest Forestry', *American Journal of Political Science*, 51 (3), 532–51.

Clemens, Elisabeth S. and James M. Cook (1999) 'Politics and Institutionalism: Explaining Durability and Change', *Annual Review of Sociology*, 25, 441–66.

Coleman, William D., Grace D. Skogstad and Michael Atkinson (1996) 'Paradigm Shifts and Policy Networks: Cumulative Change in Agriculture', *Journal of Public Policy*, 16 (3), 273–302.

Crenson, Martha A. (1971) *The Un-Politics of Air Pollution: A Study of Non-Decision Making in the Cities* (Baltimore: Johns Hopkins University Press).

Danziger, Marie (1995) 'Policy Analysis Postmodernized: Some Political and Pedagogical Ramifications', *Policy Studies Journal*, 23 (3), 435–50.

Daugbjerg, Carsten (1997) 'Policy Networks and Agricultural Policy Reforms: Explaining Deregulation in Sweden and Re-regulation in the European Community', *Governance*, 10 (2), 123–42.

Deeg, Richard (2001) *Institutional Change and the Uses and Limits of Path Dependency: The Case of German Finance* (Koln: Max-Planck-Institut für Gesellschaftsforschung).

Deeg, Richard (2005) 'Path Dependency, Institutional Complementarity and Change in National Business Systems', in G. Morgan, R. Whitley and E. Moen (eds) *Changing Capitalisms? Internationalization, Institutional Change, and Systems of Economic Organization* (Oxford: Oxford University Press,), 21–52.

deLeon, Peter (1999) 'The Stages Approach to the Policy Process: What Has It Done? Where Is It Going?', in Paul A. Sabatier (ed.) *Theories of the Policy Process* (Boulder: Westview), 19–34.

Dobuzinskis, Laurent, Michael Howlett and David Laycock (eds) (2007) *Policy Analysis in Canada: The State of the Art* (University of Toronto Press).

Doern, G. Bruce, Leslie Pal and Brian W. Tomlin (eds) (1996) *Border Crossings: The Internationalization of Canadian Public Policy* (Toronto: Oxford University Press).

Dye, Thomas R. (1972) *Understanding Public Policy* (Englewood Cliffs: Prentice-Hall).

Eichbaum, Chris and Richard Shaw (2007) 'Ministerial Advisers and the Politics of Policy-Making: Bureaucratic Permanence and Popular Control', *The Australian Journal of Public Administration*, 66 (4), 453–67.

Eldredge, Niles and Stephen Jay Gould (1972) 'Punctuated Equilibria: An Alternative to Phyletic Gradualism', in T. J. M. Schopf (ed.) *Paleobiology* (San Francisco: Freeman, Cooper) , 82–115.

Farr, James, Jacob S. Hacker and Nicole Kazee (2006) 'The Policy Scientist of Democracy: The Discipline of Harold D. Lasswell', *American Political Science Review* 100(04), 579–87.

Genschel, Phillipp (1997) 'The Dynamics of Inertia: Institutional Persistence and Change in Telecommunications and Health care', *Governance*, 10 (1), 43–66.

Hacker, Jacob S. (2004) 'Review Article: Dismantling the Health Care State? Political Institutions, Public Policies and the Comparative Politics of Health Reform', *British Journal of Political Science*, 34 (4), 693–724.

Hall, Peter A. (1989) *The Political Power of Economic Ideas: Keynesianism across Nations* (New Jersey: Princeton University).

Hall, Peter A. (1993) 'Policy Paradigms, Social Learning and the State: The Case of Economic Policy Making in Britain', *Comparative Politics*, 25 (3), 275–96.

Halligan, John (1995) 'Policy Advice and the Public Sector', in B. Guy Peters and Donald T. Savoie (eds) *Governance in a Changing Environment* (Montreal: McGill-Queen's University Press), 138–72.

Hawke, Gary Richard (1993) *Improving Policy Advice* (Wellington: Victoria University Institute of Policy Studies).

Hayes, Michael T. (1992) *Incrementalism and Public Policy* (New York: Longmans).

Howlett, Michael and Evert Lindquist (2004) 'Policy Analysis and Governance: Analytical and Policy Styles in Canada', *Journal of Comparative Policy Analysis*, 6 (3), 225–49.

Howlett, Michael and Jeremy Rayner (2006) 'Understanding the Historical Turn in the Policy Sciences: A Critique of Stochastic, Narrative, Path Dependency and Process-Sequencing Models of Policy-Making over Time', *Policy Sciences*, 39 (1), 1–18.

Howlett, Michael and M. Ramesh (2002) 'The Policy Effects of Internationalization: A Subsystem Adjustment Analysis of Policy Change', *Journal of Comparative Policy Analysis*, 4 (1), 31–50.

Howlett, Michael and M. Ramesh (2003) *Studying Public Policy: Policy Cycles and Policy Subsystems* (Toronto: Oxford University Press).

Huitt, Ralph K. (1968) 'Political Feasibility', in Austin Rannay (ed.) *Political Science and Public Policy* (Chicago: Markham), 263–76.

Hupe, Peter L. and Michael J. Hill (2006) 'The Three Actions Levels of Governance: Re-Framing the Policy Process Beyond the Stages Model', in B. Guy Peters and Jon Pierre (eds) *Handbook of Public Policy* (London: Sage), 13–30.

Jann, Werner and Kai Wegrich (2007) 'Theories of the Policy Cycle', in Frank Fischer, Gerald J. Miller and Mara S. Sidney (eds) *Handbook of Public Policy Analysis: Theory, Politics and Methods* (Boca Raton: CRC), 43–62.

Jenkins, William I. (1978) *Policy Analysis: A Political and Organizational Perspective* (London: Martin Robertson).

John, Peter and Helen Margetts (2003) 'Policy Punctuations in the UK: Fluctuations and Equilibria in Central Government Expenditure Since 1951', *Public Administration*, 81 (3), 411–32.

Jones, Charles O. (1984) *An Introduction to the Study of Public Policy* (Monterey, California: Brooks/Cole).

Kay, Adrian (2006) *The Dynamics of Public Policy: Theory and Evidence* (Cheltenham: Edward Elgar).

Kingdon, John W. (1995) *Agendas, Alternatives, and Public Policies* (Boston: Little Brown).

Kuhn, Thomas S. (1962) *The Structure of Scientific Revolutions* (University of Chicago Press).

Kuhn, T. S. (1974) 'Second Thoughts on Paradigms', in F. Suppe (ed.) *The Structure of Scientific Theories* (Urbana: University of Illinois Press), 459–82.

Kuhner, Stefan (2007) 'Country-Level Comparisons of Welfare State Change Measures: Another Facet of the Dependent Variable Problem Within the Comparative Analysis of the Welfare State', *Journal of European Social Policy*, 17 (1), 5–18.

Lasswell, Harold D. (1956) *Decision Process: Seven Categories of Functional Analysis* (College Park: University of Maryland Press).

Lasswell, Harold D. (1971) *A Pre-View of Policy Sciences* (New York: Elsevier).

Leach, William D., Neil W. Pelkey and Paul A. Sabatier (2002) 'Stakeholder Partnerships as Collaborative Policymaking: Evaluation Criteria Applied to Watershed Management in California and Washington', *Journal of Policy Analysis and Management*, 21, (4), 645–70.

Liefferink, Duncan (2006) 'The Dynamics of Policy Arrangements: Turning Round the Tetrahedron', in Bas Arts and Pieter Leroy (eds) *Institutional Dynamics in Environmental Governance* (Dordrecht: Springer), 45–68.

Lindblom, Charles E. (1959) 'The Science of Muddling Through', *Public Administration Review*, 19 (2), 79–88.

Lindner, Johann (2003) 'Institutional Stability and Change: Two Sides of the Same Coin', *Journal of European Public Policy*, 10 (6), 912–35.

Lindner, Johann and Bernard Rittberger (2003) 'The Creation, Interpretation and Contestation of Institutions - Revisiting Historical Institutionalism', *Journal of Common Market Studies*, 41 (3), 445–73.

Lindquist, E. (1998) 'A Quarter Century of Canadian Think Tanks: Evolving Institutions, Conditions and Strategies', in D. Stone, A. Denham and M. Garnett (eds) *Think Tanks Across Nations: A Comparative Approach* (Manchester University Press) 127–44.

Lyden, F. J., G. A. Shipman, R. W. Wilkinson and P. P. Le Breton. (1968) 'Decision-Flow Analysis: A Methodology for Studying the Public Policy-Making Process', in P.P. LeBreton (ed.) *Comparative Administrative Theory* (Seattle: University of Washington Press) 155–68.

Mahoney, James (2000) 'Path Dependence in Historical Sociology', *Theory and Society*, 29 (4), 507–48.

Majone, Giandomenico (1975) 'On the Notion of Political Feasibility', *European Journal of Political Research*. 3 (2) 259–74.

Maley, Maria (2000) 'Conceptualising Advisers' Policy Work: The Distinctive Policy Roles of Ministerial Advisers in the Keating Government, 1991–96', *Australian Journal of Political Science*, 35 (3), 449–49.

May, Peter J. (2005) 'Regulation and Compliance Motivations: Examining Different Approaches', *Public Administration Review*, 65 (1), 31–44.

Mayer, Igor S., Pieter W. G. Bots and Els van Daalen (2004) 'Perspectives on Policy Analysis: A Framework for Understanding and Design', *International Journal of Technology, Policy and Management*, 4 (1), 169–91.

Meltsner, Arnold J. (1972) 'Political Feasibility and Policy Analysis', *Public Administration Review*, 32 (4), 859–67.

Milner, Helen V. and Robert O. Keohane (1996) 'Internationalization and Domestic Politics: A Conclusion', in Robert O. Keohane and HelenV. Milner (eds) *Internationalization and Domestic Politics* (Cambridge University Press), 243–58.

Mortensen, Peter B. (2005) 'Policy Punctuations in Danish Local Budgeting', *Public Administration*, 83 (4), 931–50.

Peled, Alan (2002) 'Why Style Matters: A Comparison of Two Administrative Reform Initiatives in the Israeli Public Sector, 1989–1998', *Journal of Public Administration Research and Theory*, 12 (2), 217–40.

Phillips, Susan (1996) 'Discourse, Identity, and Voice: Feminist Contributions to Policy Studies', in Laurent Dobuzinskis, Michael Howlett, and David Laycock (eds) *Policy Studies in Canada: The State of the Art* (University of Toronto Press), 242–65.

Pierson, Paul (2000) 'Not Just What, but When: Timing and Sequence in Political Processes', *Studies in American Political Development*, 14 (1), 72–92.

Pierson, Paul (1993) 'When Effect Becomes Cause: Policy Feedback and Political Change', *World Politics* 45 (4), 595–628.

Prince, Michael (1983) *Policy Advice and Organizational Survival* (Aldershot: Gower).

Rochet, Claude (2004) 'Rethinking the Management of Information in the Strategic Monitoring of Public Policies by Agencies', *Industrial Management & Data Systems*, 104 (3), 201–8.

Sabatier, Paul A. (1988) 'An Advocacy Coalition Framework of Policy Change and the Role of Policy-Oriented Learning Therein', *Policy Sciences*, 21 (2), 129–68.

Sabatier, Paul A. and Hank C. Jenkins-Smith (1993) *Policy Change and Learning: An Advocacy Coalition Approach* (Boulder: Westview).

Simmons, R. H., B. W. Davis, R. J. K. Chapman and D. D. Sager (1974) 'Policy Flow Analysis: A Conceptual Model for Comparative Public Policy Research', *Western Political Quarterly*, 27 (3), 457–68.

Simon, Herbert A. (1957) *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization* (New York: Palgrave Macmillan).

Smith, Adrian (2000) 'Policy Networks and Advocacy Coalitions: Explaining Policy Change and Stability in UK Industrial Pollution Policy?', *Environment and Planning C: Government and Policy*, 18 (1), 95–114.

Smith, Richard A. (1979) 'Decision Making and Non-Decision Making in Cities: Some Implications for Community Structural Research', *American Sociological Review*, 44 (1), 147–61.

Sobeck, Joanne (2003) 'Comparing Policy Process Frameworks: What Do They Tell Us About Group Membership and Participation for Policy Development?', *Administration & Society*, 35 (3), 350–74.

Thelen, Kathleen (2003) 'How Institutions Evolve: Insights from Comparative Historical Analysis', in James Mahoney and Dietrich Rueschemeyer (eds) *Comparative Historical Analysis in the Social Sciences* (Cambridge University Press), 208–40.

Thelen, Kathleen (2004) *How Institutions Evolve: The Political Economy of Skills in Germany, Britain, the United States and Japan* (Cambridge University Press).

Tuohy, Carolyn H. (1999) *Accidental Logics: The Dynamics of Change in the Health Care Arena in the United States, Britain, and Canada* (New York: Oxford University Press).

Verschuere, Bram (2009) 'The Role of Public Agencies in the Policy Making Process', *Public Policy and Administration*, 24 (1), 23–46.

Weaver, R. Kent. and Bert. A. Rockman (1993) 'When and How do Institutions Matter?', in R. Kent Weaver and Bert A. Rockman (eds) *Do Institutions Matter? Government Capabilities in the United States and Abroad* (Washington, D.C. Brookings Institutions), 445–61

Weimer, David L., and Aidan R. Vining (1999) *Policy Analysis: Concepts and Practice* (New Jersey: Prentice Hall).

Wollmann, Hellmut (1989) 'Policy Analysis in West Germany's Federal Government: A Case of Unfinished Governmental and Administrative Modernization?', *Governance*, 2 (3), 233–66.

Yanow, Dvora (1992) 'Silences in Public Policy Discourse: Organizational and Policy Myths', *Journal of Public Administration research and Theory*, 2 (4), 399–423.

3
Selecting Cases and Inferential Types in Comparative Public Policy Research

Jeroen van der Heijden

Introduction[1]

Although many scholars claim to present comparative public policy research (see the various journals and books in this field with the word 'comparative' in their titles), only a few scholars appear to have *designed* a comparative study. My critique here echoes Benoît Rihoux' observation that, in practice, comparisons are too often 'rather loose or not formalized' (Rihoux, 2006: p. 681), and Ahrend Lijphart's observation that the comparative method appears so basic and apparently simple that when applied it often 'indicates the *how* but does not specify the *what* of the analysis' (Lijphart, 1971: p. 682). Playing the devil's advocate, I would like to take these observations a step further and claim that much of published comparative public policy research is *ex-post facto* comparative and not *a priori* comparative in nature. I understand this is a bold and sweeping statement, but in reading classic and contemporary comparative works, only a handful may be termed comparative by design (for example, Verba and Nie, 1972; Skocpol, 1979; Vogel, 1996; Levi-Faur, 2006b), whereas many works are comparative by outcome, or are studies that present various examples of a phenomenon of interest without actually having a rationale for comparatively studying these examples. Throughout this book many guidelines will be provided as to how to *design* comparative public policy research. This chapter mainly addresses the topics of case selection and the nature of the inferences drawn from comparatively analysing different cases.

In designing a comparative public policy research, the first question to be addressed is: Why compare at all? The short answer is that in studying policy the comparative method is superior to both single-N and large-N studies. It is superior to single-N studies in that the comparative method

aims to find basic patterns or tendencies in social phenomena that are likely to exist in a similar, but not identical, manner in other settings. It is superior to large-N studies as it: (i) allows to address combinations of explanatory variables in situations where the relevant number of cases is often too small for applying statistical tools; (ii) addresses irregularities ('outliers') as relevant in drawing up inferences; (iii) allows the researcher to set the boundaries of the generalizability of findings; and, (iv) forces an in-depth understanding of the cases under analysis (Ragin, 1987).

The second question to be answered is: What provides the rationale for comparing *certain* cases? I argue that answering this question starts with introducing the particular social phenomena of interest, the topic of inquiry – for example, revolutions (for example, Skocpol, 1979), voter turnout (for example, Verba and Nie, 1972) or regulatory approaches (for example, Jordana and Levi-Faur, 2004) – and ask our cases how we may better understand the phenomena of interest. For instance, we wish to know how *leadership* affects revolutions, how nations' *party systems* affect voter turnout or how *economic development* affects regulatory approaches. Based on existing literature or experiences the researcher may be able to draw up working hypothesis on how these potential explanatory variables are related to the phenomenon of inquiry. Both the presence or absence of the phenomena of interest *and* the presence or absence of these potential explanatory variables provide a rationale to *select* certain cases of others. The strength of the comparative method is that it allows the researcher to *select* her cases. This possibility to select cases allows for testing and better understanding hypothesized associations between dependent and explanatory variables.

This then brings us to the main point of this chapter: it is in carefully selecting cases that the researcher allows herself to test and better understand associations between variables. Further, as explained in this chapter, the actual choice of the cases to be analysed impacts on the type of inferential technique that may be applied, yet the type of inference impacts on the breadth and depth of the conclusions to be drawn. In short, case selection and selecting the type of inference to be made are *integrated* steps within the comparative research design. Disregarding the careful *selection* of cases and inferential techniques will, as is made clear throughout this chapter, negatively impact on the validity of the research's findings.

Different types of comparative research designs, different inferential techniques

Before addressing the topic of case selection, I first wish to address the topic of the different inferential techniques that may be pursued in

comparative research. There is no single approach to comparative research. Different research designs, however, have a different impact on the inferential technique to be used. This section discusses the two major research designs used: the most-similar-system and the most-different-system designs. In addition, and based on these two major research designs, this presents what are often considered to be the four possible inferential techniques in comparative public policy research.

Most-similar-system and most-different-system designs

Many treaties in this field refer to Mill's (1851 [1843]: Book II, Ch. 8) method of difference and method of agreement as starting point of the comparative method. Mill's *method of difference* argues that a causal association between a dependent (the social phenomenon of study) and explanatory variable (those variables that are expected to explain the phenomenon, or changes in the phenomenon of study) exists if the dependent variable is observed in only one of two settings that are fully similar except for the presence of the explanatory variable in one setting and the absence of that variable in the other. For instance, if country A and country B are fully similar except for a multiparty system in country A and a two-party system in country B and there is a strong voter turn-out during elections in country A and not in country B, we may assume a causal association between the number of parties in a political system and voter turnout during elections. Then, Mill's *method of agreement* argues that a causal association between a dependent and explanatory variable exists if the dependent variable is observed in two settings that are fully different except for the presence of the explanatory variable in both settings. For instance, if country A and B both show low voter turnout during elections and have a multiparty system, but are fully different on all other potential explanatory variables we may assume an association between multiparty systems and voter turnout (for more sophisticated insights into voter turnout, see Powell, 1982).

In short, in understanding an association between a dependent and explanatory variable, Mill challenges us to address both the control (the potential explanatory variables) *and* dependent variables. The variable associated with the phenomenon of inquiry is the one that makes the difference in the settings under analysis. It goes without saying that in political reality one is unlikely to find settings that are indeed fully similar or different except for a dependent and explanatory variable. Further, the designs are subject to severe selection bias as the cases to be studied are selected based on the dependent variable. Finally, the methods have a focus on identifying a *single* explanatory variable, whereas in social reality we most likely find various variables interacting in causing an

event – multiple conjunctural causation (for an extensive critique to Mill's methods, see Lieberson, 1992).[2]

These critiques are the starting point of an influential work in the development of the comparative method: Przeworski and Teune's *Logic of Comparative Social Enquiry* (1970). The importance of Prezworski and Teune's work lies in the fact that they aim to find a middle way between case-oriented and variable-oriented research. Instead of looking at cases (or systems) as a whole, the authors distinguish in systems and levels of systems. A system may be a sector or a nation; levels within these systems may be individuals, occupations, local communities or social classes. Systems may not only vary as a whole, they also may vary on their different levels, referred to as inter-systemic differences. In understanding an association between a dependent and explanatory variable Przeworski and Teune do not reason from a dependent variable to an explanatory variable, but from a *potential* set of explanatory variables (the system and its inter-sytem characteristics) to the dependent variable. Based on this reasoning, the authors introduce what they refer to as the most-similar-system and the most-different-system designs.[3]

The *most-similar-system design* builds on the logic that the more similar the settings being compared, the simpler it should be to trace an association between dependent and explanatory variables. Inter-systemic differences here explain similarities or differences in the phenomenon under study. This design helps the researcher to reduce the number of dependent and explanatory variables to study under controlled conditions – that is, the cases are selected. The design is often considered an adaptation of Mill's method of difference, but strongly differs from it (for an extensive discussion, see Levi-Faur, 2006a). Where Mill's design selects cases based on dependent and explanatory variables, Przeworski and Teune's design selects cases based on explanatory variables only – which may overcome the earlier mentioned selection bias related to Mill's method of difference. Yet, the design is prone to similar criticism. In reality it will be difficult to find settings that meet this design's requirements (for similarity, see also Przeworski and Teune, 1970; Lijphart, 1971). Besides this the design suffers from its own selection bias: when selecting cases based on similarities one is most likely to find that the (small number of) differences between the cases explain potential differences in the phenomenon of study. Increasing the number of cases does not bring in the possibility to test whether other variables may explain this variation (see Meckstroth, 1975). Also, the more similar the settings under analysis, the less strong the arguments the researcher can make on the generalizability, or external validity, of her study.

The *most-different-system design* aims to provide an answer to exactly these problems. The logic is to compare cases that are as different as possible, demonstrating the robustness of an association between dependent and explanatory variables. Such a design assumes that the argument of the research is better supported by demonstrating that the observed relationship holds despite the wide range of contrasting cases. Further, this approach starts with considering within-system characteristics as relevant variables that may falsify an association between a dependent and an explanatory variable. For instance, when addressing the question of voter turnout the researcher's initial focus would not be on multi-party systems versus two-party systems (system characteristic) but, for instance, on differences in labour organizations, male-female ratios and income amongst countries (within-system characteristics). If the within-system characteristics are similar (or vary in a similar manner) in a set of cases and the phenomenon of interest is similar amongst settings (for example, voter turnout), *but* the system characteristics are different (for example, multi-party system versus two-party system) then the researcher eliminates these system characteristics as associated with the dependent variable. To Przeworski and Teune the advantage of the most-different-system design is that it enables one to '[formulate] statements that are valid regardless of the systems within the observation is made' (Przeworski and Teune, 1970: p. 39). The logic of this approach relates to what may be termed disconfirming evidence (for example, Creswell and Miller, 2000) – instead of providing numerical evidence of an association by studying similar cases, the researcher aims to provide evidence by studying variance in cases.

Understanding the social phenomenon of study: Different inferential techniques

None of the above designs is satisfying in providing a logical and undisputable approach to inferring from explanatory variables to the dependent variable (the social phenomena that is addressed in the study). Inference is a gradual and cumulative process of elimination, falsification, discovery and corroboration – though such 'muddy' processes are prone to criticism. The logic of comparative analysis, however, allows for stating *a priori* assumptions on the inferences drawn. Introducing such assumptions may provide more clarity and transparency to the inferential process. Combining the designs by Mill (1851 [1843]) and Przeworski and Teune (1970) provides four inferential techniques that may help to strengthen the process of inference in comparative public policy research (Levi-Faur, 2006a). Further, these four techniques address the four natural types, or

Table 3.1 A priori assumptions for different inferential techniques (based on Levi-Faur, 2004)

	Mill's method of difference	**Mill's method of agreement**
Most-similar-system design	*Technique 1*: dealing with differences in similar cases. *Assumption*: The few differences in explanatory variables are associated with the differences in the dependent variable.	*Technique 2*: dealing with similarities in similar cases. *Assumption*: It is unlikely that the few differences in explanatory variables are associated with the similarities in the dependent variable.
Most-different-system-design	*Technique 3*: dealing with differences in different cases. *Assumption*: It is unlikely that the few similarities in explanatory variables are associated with the variance in the dependent variable.	*Technique 4*: dealing with similarities in different cases. *Assumption*: It is unlikely that the differences in explanatory variables are associated with the similarities in the dependent variable.

settings, of potential pairwise case comparisons the researcher may face: first, most-similar cases with different outcomes; second, most-similar cases with similar outcomes; third, most-different cases with different outcomes; and, fourth, most-different cases with similar outcomes. Table 3.1 provides an overview of the *a priori* assumptions per inferential technique.

The first technique combines the logic of the most-similar-system design and Mill's method of difference. This technique may be applied to a setting of two or more cases that show minimal variance in the control variables and maximal variance in the dependent variable. The few differences in the control variables then are assumed explanatory for, or at least associated with, the difference in the dependent variable. An (oversimplified) example would be two countries (A and B) that are similar on a number of characteristics that are hypothesized as related to voter turnout during elections (for example, both have a two-party system, flourishing economies, accessible and understandable voting processes and so on.), but show significant variance in voter turnout. Here the researcher may address other potential explanatory variables and may, for instance, find that in country A the weather was warm and sunny during elections, but cold and rainy in country B. If, in this hypothetical case, weather conditions were among the few relevant differing

explanatory variables, the researcher may assume an association between weather on the day of elections and voter turnout.

The second technique combines the logic of the most-similar-system design and Mill's method of agreement. This technique may be applied to a setting of two or more cases that show minimal variance in the control variables and minimal variance in the dependent variable. The few differences in the control variables are then considered as less likely related with the similarity in the dependent variable. Using the same example, but now with similar patterns in voter turnout, the researcher may assume that the difference in weather conditions is not associated with voter turnout in countries A and B.

The third technique combines the logic of the most-different-system design and Mill's method of difference. This technique may be applied to a setting of two or more cases that show maximal variance in the control variables and maximal variance in dependent variables. The few similarities in control variables are then considered as less likely related with the variance in the dependent variable. An example would be two countries (A and B) that are highly different in characteristics that are hypothesized as related to voter turnout during elections (for example, country A is characterized by a two-party system, a flourishing economy and accessible and understandable voting processes; and country B is characterized by a multi-party system, a troublesome economy and a complex voting process) and show significant variance in voter turnout. Here the researcher may address other potential explanatory variables and may, for instance, find that in both countries the weather was warm and sunny during elections. Would, in this hypothetical case, weather conditions be among the few relevant similar explanatory variables; the researcher may assume there is no association between weather on the day of elections and voter turnout.

The fourth technique combines the logic of the most-different-system design and Mill's method of agreement. This technique may be applied to a setting of two or more cases that show maximal variance in the control variables and minimal variance in the dependent variables. The differences in control variables are then considered as less likely related with the similarities in the dependent variable. Using the same example as under the third technique, but now with similar patterns in voter turnout, the researcher's hypotheses on the characteristics associated with voter turnout are falsified. Note that this technique does not allow for statements on a potential association between weather conditions and voter turnout – the dependent variables under analysis are kept to a minimum.

To give an example, if the researcher introduces a hypothesis that voter turnout is affected by the type of party system in a country she may wish to test this hypothesis by either inferential techniques 1, 3 or 4, but each technique would allow for different conclusions. Technique 1 would provide mild evidence to reject her hypothesis (that is, similarity in party system and difference in voter turnout would at least bring evidence that the level of voter turnout is not likely to be associated with that particular type of party system, but this conclusion may not be extrapolated to other party systems); technique 3 would provide mild evidence to support her hypothesis (that is, difference in party system and difference in voter turnout are in line with the hypothesis, but this conclusion may not be generalized beyond the cases studied); and, technique 4 would provide strong evidence to reject her hypothesis (that is, difference in party system and similarity in voter turnout fully rejects the hypothesis drawn).

Technique 2 is of limited avail in this example as it would provide mild evidence that a particular type of party system is likely associated with a particular level of voter turnout, but does not address the hypothesis drawn. Not only does the research design affect the inferences that may be drawn, it is in carefully selecting cases that the researcher can apply a certain inferential technique.

Selecting cases in comparative public policy research

As has become clear from the previous section, the type of comparative research design chosen informs the inferential technique that may be used, and this inferential technique impacts on the conclusions the researcher may reach. The same holds for case selection: the cases selected ultimately set boundaries to the conclusions the researcher may draw. Two questions are of relevance in selecting cases. The first question, 'What makes a case?', relates to the discussion in the previous paragraph. It is in the cases selected that the researcher finds data to support or reject her hypothesis. The second question, 'How many cases are needed?', relates to the certainty as to which she can draw conclusions. A long-standing critique to the comparative method is what Arend Lijphart has termed the principal problem of the comparative method: 'many variables, small number of cases' (Lijphart, 1971: p. 685) – 'many-V/small-N'. In short, so Lijphart's argument goes, if one aims to find the factor(s) that are associated with the social phenomenon under study (an association between dependent and explanatory variables), one quickly runs out of cases when the number of potential explanatory variables is too large. These two questions are the topic of this section.

What makes a case?

It is all too easy to surpass this seemingly obvious question when designing a comparative public policy research. Looking at the literature, we find no easy answer to this question (for example, Ragin and Becker, 1992; Yin, 2003; Collier et al., 2004; Levy, 2008; VanWynsberghe and Khan, 2008). In the still topical edited volume *What is a Case?* (Ragin and Becker, 1992), Charles Ragin concludes that it is of little avail to come to a finite answer to this question. Casing, so argues Ragin (Ragin and Becker, 1992: p. 218), 'can bring operational closure to some problematic relationship between ideas and evidence, between theory and data'. Following Ragin's work, we look upon a case as a real-world occurrence of a social phenomenon – for example, revolutions (for example, Skocpol, 1979), democracy (for example, Verba and Nie, 1972), social capital (Putnam et al., 1993) or regulatory approaches (for example, Jordana and Levi-Faur, 2004).

To gain insight in the case under analysis its context needs to be studied (see VanWynsberghe and Khan, 2008); for instance, the level of democracy *in* a country; the behaviour of firms *in* a policy sector; social capital *in* a certain period of time. Here the country, policy sector and time period provide the set of potentially explanatory variables. Depending on the phenomenon of study, smaller case contexts may be firms, professional associations or even individuals (this relates to the earlier discussed work of Przeworski and Teune, 1970; also, Singer, 1961).[4]

But where to get cases from? Traditional and contemporary comparative research designs tend to understand cases, and certain social phenomena, by addressing these in specific sets of contexts. Four approaches stand out (Levi-Faur, 2004). First, '[t]he *national patterns approach* suggests that national-level characteristics exert a major impact on policy, politics, economics and society' (Levi-Faur, 2004: p. 180, my emphasis). Such characteristics may be (Van Waarden, 1995): the national form of intermediation between state and society; preferences for action; the policies' extent for integration; flexibility of rule formulation and application; state–clientele relations and network social relations. Second, somewhat opposing the national patterns approach the *policy sector approach* may be summarized in two propositions: '[First] that the style of policy making and the nature of policy conflicts will vary significantly from sector to sector. ... [And second] that policy making in a particular sector will exhibit strong similarities, whatever its national context' (Freeman, 1985: p. 486). Here explanatory variables may be found in particular changes in technology, market structure or regulatory approaches in policy sectors (Hollingsworth et al., 1994). Third, a different point of view is taken within the *international regimes approach*.

Here it is argued that international regulatory regimes result in regulatory reforms. Different international regimes may result in different reforms. Explanatory variables may be found in: key actors within an international regime, governing principles of the regime and mechanisms of globalization (Young, 1999; Braithwaite and Drahos, 2000). Fourth and final, the *temporal patterns approach* looks at old and new states of affairs within, for instance, countries, policy sectors or international regimes. By treating different 'eras' between these states of affairs researchers aim to find similarities and differences in these periods of policy or regulatory change.

In my opinion hypotheses drawn on assumed associations between dependent and explanatory variables should inform the selection of case contexts, but researchers following a less deductive paradigm may hold other options. Further, the specific approach used will generate a certain range of findings, whilst excluding others – a selection bias. For instance, when aiming one's research lens at country patterns only and not on potential differences between policy sectors, one is most likely to trace an association between variance in country characteristics and variance in policy or regulatory outcomes in these countries irrespective of the particular policy area studied. Combining the various approaches – to the extent possible – may help to overcome the shortcomings of the approaches whilst at the same time playing out their strengths – this is discussed further below (under the heading 'compound comparative research design').

How many cases are needed?

Carefully selecting cases may help the researcher to build strong claims on the validity of the conclusions drawn. She will however still face Lijphart's 'many-V/small-N' problem. In other words, how many cases need to be studied in order to understand the strength of the causal associations drawn? After all, given the wide range of potential explanatory variables, the researcher may perhaps choose a set of cases that holds some, but not all, necessary variables that are needed for the social phenomena to occur. The more relevant variables that can be eliminated as to *not* be associated with the social phenomena of interest, or the more variables that can be evidenced as to *be* associated, the stronger the claims to be made on the validity of the findings.

When reviewing the literature on comparative research designs one finds, again, that there is no easy answer to the question of how many cases are needed – if there is an answer at all (for example, Lijphart, 1971; Ragin, 1987; Peters, 1998; Brady and Collier, 2004; Della Porta and

Keating, 2008; Hopkin, 2010). It can be argued that when the number of cases increases, the researcher is better able to understand and explain: (i) the range of explanatory variables that affect the outcome (the spread of the causal association); and (ii) the strength or impact of the causal association (cf. Franklin, 2008). At the same time, increasing the number of cases most likely implies sacrificing the depth of the study (cf. Peters, 1998), losing exactly the strengths of case-oriented research. To quote Gerring (2007: p. 348): 'Research designs invariably face a choice between knowing more about less and knowing less about more.'

It goes without saying that a comparative research design requires cases to *compare*. This makes it relative easy to set a lower limit for a research design – at least two cases.[5] A 'good' upper limit is more difficult to define. Yet, from the above discussion on most-similar-system designs and most-different-system designs we have learned that the validity of research findings largely depends on whether and how extraneous variance is *controlled* for. We further learned that a comparative research design requests for a careful *selection* of cases. These two lessons may help us in deciding how many cases we wish to study.

First, the comparative method differs from experiments and the statistical method in that it deliberately selects and matches settings to be able to detect the presence or absence of such association between dependent and explanatory variables (Frendreis, 1983). This need to carefully select the cases inevitably brings a selection bias into the study – cases will partly be *selected* on the presence of the dependent variable (see Geddes, 1990; King et al., 1994). Being able to study this dependent variable in depth partly is the strength of case-oriented research (Bennett and Elman, 2006). Theoretically the problems of selection bias can be overcome by selecting cases based on explanatory variables and not dependent variables (Geddes, 1990; Brady and Collier, 2004). Yet, contrary to statistical analysis it is unlikely that a comparative research design relies on a random sampling of cases. As such there is no need to increase the number of cases to provide evidence that a certain causal association between dependent and explanatory variables is (statistically) significant or not – there is no random sampling in (most) comparative-oriented research designs.

Second, bringing in more cases into the study inevitably implies bringing in more extraneous variance into most-similar-system designs, and less extraneous variance into most-different-system designs. After all, in the hypothetical example of two settings under analysis being fully different on all but one explanatory variable, bringing in a third setting increases the possibility that it will show some characteristics

from either one or both of the initial settings under analysis and thus lowers extraneous variance. This reasoning would result in a preference for a paired-comparative research design – that is, comparatively studying two cases. Yet, as we have seen from the discussion on the most-similar and most-different-system designs, these designs fall short exactly in their inability to control for the full absence or full presence of extraneous variance. The smaller the number of cases under analysis the more likely this inability goes unnoticed, which may result in assumed causal associations between dependent and explanatory variables where in fact more variables explain the occurrence of the dependent variable – a faulty interference (also, Geddes, 1990).

   Thus, if studying two cases only may very well provide too little external validity but adding cases brings in problems of controlling for extraneous variance, what is it that the researcher should do? From the above discussion it has become clear that preferably cases are added to the research design to increase the external validity, or generalizability, of findings without changing the original extraneous variance. This may be done not so much by focussing on the actual *number* of cases, but by focussing on the actual *types* of case environments (cf. Levi-Faur, 2004; Della Porta, 2008). Phrased differently, instead of aiming to strengthen a research by a sheer increase in the number of cases, the researcher may better aim to strengthen her research by increasing the number of potential explanatory variables.

Compound comparative research design

The number of potentially explanatory variables may very well be increased if the researcher chooses to study a similar case in different contexts – for instance, revolutions in different countries and revolutions in different industries. Various combinations of the four different case contexts discussed before can be made. An evident combination is to combine the national and the sectoral approach. Vogel's *Freer Markets, More Rules* (1996) provides an exemplary example by comparatively analysing (initially) the regulation of telecommunications and financial sectors in Japan and the United Kingdom. In combining the national and the sectoral approach the researcher may question whether the generic characteristics of a chosen sector and nation-specific characteristics are associated with a certain outcome – for example, technological, economic and political characteristics of the sector *and* the nation. A further question in this particular setup is whether certain specific [nation*sector] characteristics affect the phenomenon of study (Levi-Faur, 2006c). Combining the national and temporal approach is another

possible research design. Putnam et al.'s *Making Democracy Work* (1993) provides an intriguing example by comparatively analysing a number of regional Italian governments over time. Although Italy is a single country Putnam et al. presents a narrative on how inter-country differences affect development of institutions. In combining the (sub)national and temporal patterns approach the researcher may question whether generic characteristics of chosen time periods and nation-specific (or in Putnam et al.'s case, region-specific) characteristics are associated with a certain outcome. A further question in this particular setup is whether certain specific [nation*time] characteristics affect the phenomenon of study (also, Lynggaard, 2011).

Combining different approaches helps researchers to gain a better understanding of the association(s) between independent and explanatory variables under study. Further, the compound research designs help the researcher to increase the number of possible pairwise comparisons by increasing the number of cases or units of analysis. If we, for example, look at the outcomes of contemporary regulatory policies in two sectors in two countries (for example, sector X and sector Y, country A and country B) we are able to conduct six pairwise comparisons: (i) country A and country B; (ii) sector X and sector Y; (iii) country A*sector X and country B*sector X; (iv) country A*sector Y and country B*sector Y; (v) country A*sector X and country B*sector Y; and, (vi) country A*sector Y and country B*sector X. A further step would be to include a temporal aspect into the analysis by comparing the contemporary regulatory policies with former regulatory policies. This provides for 21 different pairwise comparisons (Levi-Faur, 2006c: p. 376).

Besides increasing the number of possible pairwise comparisons and so increasing the validity of findings, a further advantage of a compound research design lies in the fact that adding a case or a unit of analysis implies a relatively limited increase of extraneous variance. After all, when initially comparing a phenomenon in two countries (A and B) and adding a third (C), the extraneous variance of that third country may strongly affect the internal validity of the study, whereas only one or two extra comparisons are added (A and C; B and C). The extraneous variance of that third country needs to be studied to understand its impact. Yet, when initially comparing a phenomenon in a sector in two countries and enriching the study by addressing that phenomenon in another sector in both countries, the internal validity of the study may be strengthened as the respective contexts are studied in more depth *and* five extra comparisons are added (see above). Also, here the extraneous variance brought in by the sectoral comparison needs to be

studied. However, we can assume that studying the same phenomenon in different sectoral contexts, but within the same country context, offers a better solution to the 'many-V/small-N' problem than studying the same unit in an ever increasing range of contexts (see Lijphart, 1971: p. 64). Phrased differently, the relative increase of extraneous variance by adding units within the larger context of the compound research design is smaller than adding units *and* their contexts to most-similar-system or most-different-system designs. In addition, for practical reasons the researcher may wish to combine, for example, the national patterns approach with the policy sectors approach as the study of another country is too costly, or another country is simply not available (for example, Frendreis, 1983).

Thus, by applying a compound research design the researcher is able to exploit the advantages of comparative research: adding an additional case does provide the researcher the opportunity to both question whether an assumed association between dependent and explanatory variables holds in similar settings (for example, a range of developed countries), *and* whether the association holds in different settings (for example, the financial sector and the health and safety sector). The compound research design allows for the testing of a wide range of (potential) explanatory variables at different systemic levels of different case-contexts. At the same time, by keeping the increase in extraneous variance relatively low, the researcher is able to present more valid and more generalizable findings and partly overcome the selection bias that is inherent to comparative research designs.

From ideal type designs to the reality of comparative public policy research

The earlier sections have provided insight into an 'ideal-type' comparative public policy research design. In practice, however, the researcher will face constraints that hamper her to follow this ideal-type design, which most likely have to do with limitations to funds or time needed to follow this research design. Besides such constraints, other situations may obstruct following the ideal-type research design. Cases that allow for sufficient testing of a wide range of explanatory variables may not all be accessible. For instance, in studying revolutions a researcher may find that a country currently at war would constitute a good case to include in the research design while the actual undertaking of research in that country may be too dangerous. Such constraints impact the validity of the research findings, and on the ability to move from in-depth understanding of the

cases under analysis to generalizability of findings. This section addresses these two specific problems.

How to strengthen the internal and external validity of the study when facing limited means?

Internal validity may be understood as 'the degree to which descriptive or causal inferences from a given set of cases are correct for those cases' (Brady and Collier, 2004: p. 292). The strength of the case-study approach in comparative analysis is that cases and, related, their contexts can be studied in depth, thus helping the researcher to strengthen the internal validity of her study. Various treaties on case-study methodology discuss the problem of internal validity of case studies, with process tracing, pattern matching and causal narratives being dominant and accepted strategies to deal with this problem (Mahoney, 2000; Bennett and Elman, 2006; Venesson, 2008).

External validity may be understood as 'the degree to which descriptive or causal inferences for a given set of cases can be generalized to other cases' (Brady and Collier, 2004: p. 288). Often it is argued that internal validity and external validity are in conflict: the fewer cases or units studied, the smaller the external validity of findings as cases are comparable on a small number of variables only; the more cases or units studied, the smaller the internal validity as the researcher is most likely unable to study all cases or units in depth. Yet, is this conflict a problem for the comparative method? In understanding the *extent* to which a comparative research can produce generalizable findings I argue it is not a problem.

The researcher applying the comparative method should however be aware of the extent to which her findings may be generalized. In other words, what is it the comparative method can do, and what is it the comparative method cannot do? What the comparative method as discussed in this chapter cannot produce is empirical generalization. Simply for the fact that it is impossible to meet the *ceteris paribus* condition – that is, one cannot control all extraneous variance (cf. Lijphart, 1971; Payne and Williams, 2005). As such it is impossible to make statistical inferences and calculate probabilistic estimates for a wider universe of cases or units. But, there is a distinction between generalizing findings to populations and generalizing findings to theoretical propositions (Yin, 2003). And the latter is what the comparative method *can* produce: analytical or moderatum generalizations, or, in certain instances, theoretical generalizations (cf. Payne and Williams, 2005). *Moderatum generalizations* do not attempt to 'produce sweeping sociological statements that hold good over long periods of time, or across ranges of cultures' but

provide bounded 'claims to basic patterns, or tendencies, so that other studies are likely to find something similar but not identical' (Payne and Williams, 2005: p. 297, 306). *Theoretical generalizations* may be produced if the comparative study aims to falsify an existing theory by comparing non-typical, critical or extreme cases (cf. Yin, 2003; Payne and Williams, 2005).[6]

Keeping these limitations in mind, our concern as comparativists should not be to maximize external validity at the cost of internal validity, but to provide a level of external validity that suits the level and type of generalization we pursue. A stepwise approach to the comparative study by a controlled increasing of cases under analysis may help to do so.

A stepwise approach to comparative public policy research

Taking the attractiveness of the comparative method as point of departure – in-depth understanding *and* generalizability – we can rightfully question how to make the move from in-depth knowledge to generalizable findings when facing limitations of means. In addition to the earlier discussed approaches (the combining of the most-similar-system and most-different-system designs, and the compound research design) an additional approach is a stepwise increase of cases under analysis (Vogel, 1996; Levi-Faur, 2004) – keeping in mind that the case is the phenomenon of interest, not the context within which it is studied. The immediate problem the researcher faces, however, is whether to choose a most-similar-system or most-different-system approach. Partly this question may be answered given the goal of the research. The comparative method can be applied for developing and testing hypothesis (cf. Meckstroth, 1975). The 'many-V/small-N' problem appears most problematic when carrying out hypothesis-testing research. After all, when carrying out hypothesis-developing research one may argue that the more potential explanatory variables the cases' contexts provide the better, since these can be eliminated through falsification.

A hypothesis-testing research may be addressed by using both the most-similar-system and the most-different-system designs. I argue that, in such research, the researcher should not be hamstrung by the question whether to choose a most-similar-system or most-different-system approach. The first step in a hypothesis-testing comparative analysis is to take the existing literature and theorizing as point of departure: What relationships between dependent and explanatory variables are hypothesized, and why? If the researcher's expectation is that system characteristics (for example, party system) explain a certain phenomenon (for example, voter turnout) either a most-similar-system design or a most-different-system design can be applied. If the researcher's expectation is

that inter-system characteristics (for example, communities, social class) explain a certain outcome (for example, voter turnout) a most-different-system design appears more suitable (cf. Przeworski and Teune, 1970).

The second step is to select a number of primary cases and study these in depth. The selection of the primary cases or units logically follows from the hypothesis developed. In studying the primary cases the researcher aims to falsify the hypotheses developed. The researcher first analyses her cases using either the most-similar-systems or the most-different-systems approach, and then investigates a different set of cases using the contrasting approach. Thus, if the starting point of the research was a most-similar-systems design, the researcher now analyses whether the cases provide for relevant differences that impact on the phenomenon under study. And, if the starting point of the research was a most-different-systems design, the researcher now analyses whether the cases provide for relevant similarities that impact on the phenomenon under study.[7] By combining the two research designs the researcher will not only be able to test whether the initial hypotheses can be falsified, she is also able to draw up alternative hypotheses. As such this second step combines both abductive and deductive logic (Aliseda, 2004; Paavola, 2004). It allows for a structured and inductive testing of hypothesis, whilst being open to abductively 'discovering' new hypothesis that may be tested in a later stage of the research – hence strengthening the study's internal validity.

To give an example, the first step of a hypothetical research design comparatively analyses voluntary environmental governance arrangements in a set of countries (Van der Heijden, 2012). In doing so it addresses the development of non-mandatory sustainable building certification regimes in Germany and Australia and questions whether country characteristics (that is, existing environmental regulation, market conditions, government–industry relationships and so on) affect the development and performance of such regimes. Note that the researcher studies two cases (development and performance) in a range of units (certification regimes).

The third step is to select a number of secondary cases *within the primary case contexts* and study these on those variables that had some explanatory power in the second step of the analysis. Thus, here the researcher can suffice to study these secondary cases in less depth than the primary cases. In this step the researcher aims to understand whether the non-falsified *and* potentially newly traced association(s) between explanatory and dependent variables holds in a larger number of cases, but within the primary cases' contexts. This step helps the researcher to generalize

findings beyond the primary cases studied, but within the contexts of the initial cases selected. The earlier discussed compound research design may help the researcher to strengthen the external validity in this stage of the study, and allow her to draw moderatum generalizations. To continue our example, the researcher now addresses another policy area to the hypothetical study. She chooses water management and selects a number of units to study her cases, for instance, sustainable water management certification regimes. Here we see an example of a [country*sector] compound design.

Potential additional steps are to select a number of tertiary, and so on, cases *outside the primary cases contexts* and study these on those variables that had some explanatory power in the third step. Again the researcher can study these tertiary, and so on, cases in less depth than the primary and secondary cases or units as the aim of this step is to understand whether earlier non-falsified associations hold in a larger number of cases, by contexts outside the primary cases. The advantage of this last step is that it enhances the external validity of findings, but it weakens the internal validity of the study as the new case contexts bring a new set of extraneous variables into the study – that is, potential explanatory variables. In our example, the researcher would add other countries to the study (new case contexts), and include a temporal aspect to the study by addressing her cases in voluntary environmental governance in different time areas, or she could even address her cases in different international regime settings.

Conclusion

This chapter addressed the importance of careful selection of cases and inferential techniques when carrying out comparative public policy research. Between the lines this chapter presented what I think is an ideal-type research design for comparative public policy research. In summary this ideal-type research design would follow these facets: (i) choose a social phenomenon to study; (ii) develop hypotheses on how different variables affect this social phenomenon; (iii) choose an approach as to how to test these hypotheses and state what assumptions may be made from this approach – that is, choose a type of inference; (iv) carefully select a pair of cases within similar case contexts that allows the application of the relevant inferential techniques – that is, the phenomenon of inquiry in different countries, sectors, eras or international regimes; (v) increase the number of potential explanatory variables by including cases from different case contexts following a compound research approach; (vi) repeat the fifth step until the research findings

allow for sufficient validity to draw conclusions; and (vii) report findings, be explicit in the research design and approach chosen and discuss the limitation to the internal and external validity of findings. I am aware of the constraints of real-world comparative public policy research, especially as to facets iv, v and vi of this ideal-type research design. To overcome such constraints I have discussed a number of hands-on strategies.

In discussing the ideal-type approach to comparative research the chapter addressed the oft-heard critique to comparative research – that, in practice, comparisons are too often 'rather loose or not formalized' (Rihoux, 2006: p. 681). I do not, however, claim that the approach presented here is the only suitable approach for comparative research. My aim was to present an *a priori* defined systematic approach to help the researcher in making decisions on her research *and* to help her in explaining and giving account to her readership on what has been done. First, the four inferential techniques help to stress why specific conclusions may be drawn from certain types of pairwise comparisons. Second, the balancing of internal and external validity helps in understanding what level of generalizations may be produced and at what cost. Third, the compound research design enables the researcher to find whether assumed associations between dependent and explanatory variables hold in a wide range of settings. Fourth, the controlled and stepwise increase of the number of cases under analysis helps to strengthen the external validity of the study without sacrificing its internal validity. Finally, by applying an *a priori* and formalized research design the researcher increases the transparency of the inferential process of her study.

Notes

1   I am grateful to David Levi-Faur in commenting on an earlier version of this chapter. As the reader will find, the chapter draws heavily on David Levi-Faur's work. This is done with his full permission. I am further grateful for helpful comments and advise from the Drafts-R-Us Discussion Group at RegNet and the editors of this volume. That said, all errors and omissions in this chapter are my responsibility.
2   In teaching this subject I use the illustration of a fruit salad. The method of difference applies when we have a salad that fully consists of bananas and one that has one strawberry in it. If we find that one of my students does not like the latter salad, then, according to Mill, this must be because she does not like strawberries – for example the strawberry is the explanatory variable here. The method of agreement, then, would argue that we have two fruit salads. One is made out of a banana, some strawberries, some raspberries and a pineapple; the other is made up out of a mango, a passion fruit, some raspberries and some blackberries. If one of my students likes *both* the fruit salads, then, according to Mill, this must be because he is fond of raspberries.

3 Getting back to the fruit salad example, fruit may be broken down into categories such as 'fibres', 'percentage of sugar', 'colour', 'freshness' and so on. If in the first example of the banana–strawberry salad one of the bananas happens to be mouldy, whereas the full-banana salad is fresh, we suddenly have a different insight into why my student does not like the banana–strawberry salad – that is, it is not the strawberry that explains, but the freshness of the fruit. Also, if we find that in the other examples my student is known to have a sweet-tooth, his preference for sweet fruit in general may explain why he likes both salads – that is, not the raspberries, but the 'percentage of sugar' explains his preference.

4 Note: a single case – the social phenomenon of interest – may occur in different case contexts; whilst a single context may hold different cases – for example revolution in a country, democracy in that country, voter turnout in that country (further Gerring, 2007).

5 Though, see Tarrow's (2010) discussion on the strengths of paired-case studies compared to case studies consisting of more than two cases; and, see Gerring (2007) on the strength of single case studies – but note that these single case studies ultimately *compare* the case under analysis with existing literature or theorizing.

6 For an excellent discussion of different types of cases that may be pursued when aiming at theoretical generalizations, see the work of Seawright and Gerring (2008).

7 Contrary to note 1, this goes against the work of David Levi-Faur, who argues for a sequence of a most–similar-systems followed by a most-different-systems design. I do not see a *theoretical* rationale for that particular sequence.

Bibliography

Aliseda, Atocha (2004) 'Logics in Scientific Discovery', *Foundations of Science,* 9 (3), 339–63.

Bennett, Andrew and Elman, Colin (2006) 'Complex Causal Relations and Case Study Methods: The Example of Path Dependence', *Political Analysis,* 14 (3), 250–67.

Brady, Henry and Collier, David (eds) (2004) *Rethinking Social Inquiry Diverse Tools, Shared Standards* (Lanham: Rowman and Littlefield).

Braithwaite, John and Drahos, Peter (2000) *Global Business Regulation* (Cambridge University Press).

Collier, David, Brady, Henry and Seawright, Jason (2004) 'Sources of Leverage in Causal Inference', in Henry Brady and David Collier (eds) *Rethinking Social Inquiry* (Lanham: Rowman and Littlefield), 229–66.

Creswell, John and Miller, Dana (2000) 'Determining Validity in Qualitative Inquiry', *Theory Into Practice,* 39(3), 124–30.

Della Porta, Donatella (2008) 'Comparative Analysis: Case-Oriented Versus Variable Oriented Research', in Donatella Della Porta and Michael Keating (eds) *Approaches and Methodologies in the Social Sciences: A Pluralist Perspective* (Cambridge University Press), 198–222.

Della Porta, Donatella and Keating, Michael (eds) (2008) *Approches and Methodologies in the Social Sciences. A Pluralist Perspective* (Cambridge University Press).

Franklin, Mark (2008) 'Quantitative Analysis', in Donatella Della Porta and Michael Keating (eds) *Approaches and Methdologies in the Social Sciences: A Pluralist Perspective* (Cambridge University Press), 240–62.

Freeman, G. P. (1985) 'National Styles and Policy Sectors: Explaining Structured Variation', *Journal of Public Policy,* 5 (4), 467–96.

Frendreis, John (1983) 'Explanation of Variation and Detection of Covariation: The Purpose and Logic of Comparative Analysis', *Comparative Political Studies Comparative Political Studies,* 16 (2), 255–72.

Geddes, Barbara (1990) 'How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics', *Policy Analysis,* 2 (1), 131–50.

Gerring, John (2007) 'Is There a (Viable) Crucial-Case Method?', *Comparative Political Studies,* 40 (3), 231–53.

Hollingsworth, Roger, Schmitter, Phillipe and Streeck, Wolfgang (1994) 'Capitalism, Sectors, Institutions and Performance', in Roger Hollingsworth, Phillipe Schmitter and Wolfgang Streeck (eds) *Governing Capitalist Economies: Performance and Control of Economic Sectors* (Oxford University Press), 3–16.

Hopkin, Jonathan (2010) 'The Comparative Method', in David Marsh and Gerry Stoker (eds) *Theory and Methods in Political Science* (New York: Palgrave Macmillan), 285–307.

Jordana, Jacint and Levi-Faur, David (2004). *The Politics of Regulation: Institutions and Regulatory Reforms for the Age of Governance* (Cheltenham: Edward Elgar).

King, Gary, Keohane, Robert and Verba, Sidney (1994). *Designing Social Inquiry. Scientific Inference in Qualitative Research* (Princeton University Press).

Levi-Faur, David (2004) 'Comparative Research Designs in the Study of Regulation: How to Increase the Number of Cases Without Compromising the Strengths of Case-Oriented Analysis', in Jacint Jordana and David Levi-Faur (eds) *The Politics of Regulation. Institutions and Regulatory Reforms for the Age of Governance* (Cheltenham: Edward Elgar), 177–99.

Levi-Faur, David (2006a) 'A Question of Size?', in Bernhard Rihoux and Heike Grimm (eds) *Innovative Comparative Methods for Policy Analysis* (New York: Springer and Kluwer), 43–66.

Levi-Faur, David (2006b) 'Regulatory Capitalism: The Dynamics of Change beyond Telecoms and Electricity', *Governance,* 19 (3), 497–525.

Levi-Faur, David (2006c) 'Varieties of Regulatory Capitalism: Getting the Most Out of the Comparative Method', *Governance*, 19 (3), 367–82.

Levy, Jack (2008) 'Case Studies: Types, Designs, and Logics of Inference', *Conflict Management and Peace Science*, 25 (1), 1–18.

Lieberson, Stanley (1992) 'Small N's and Big Conclusions', in Charles Ragin and Garry Becker (eds) *What is a Case?* (Cambridge University Press), 105–18.

Lijphart, Ahrend (1971) 'Comparative Politics and the Comparative Method', *The American Political Science Review,* 65 (3), 682–93.

Lynggaard, Kennet (2011) 'Domestic Change in the Face of European Integration and Globalization: Methodological Pitfalls and Pathways', *Comparative European Politics,* 9 (1), 18–37.

Mahoney, James (2000) 'Strategies of Causal Inference in Small-N Analysis', *Sociological Methods Research,* 28 (4), 387–424.

Meckstroth, Theodore (1975) '"Most Different Systems" and "Most Similar Systems": A Study in the Logic of Comparative Enquiry', *Comparative Political Studies,* 8 (2), 132–57.

Mill, John Stuart (1851) *A System of Logic,* 3rd edn (London: Longmans, Green, and Co.).

Paavola, Sami (2004) 'Abduction as a Logic and Methodology of Discovery: The Importance of Strategies', *Foundations of Science,* 9 (3), 267–83.

Payne, Geoff and Williams, Malcolm (2005) 'Generalization in Qualitative Research', *Sociology,* 39 (2), 295–314.

Peters, Guy (1998) *Comparative Politics: Theory and Methods* (Houndmills: Palgrave Macmillan).

Powell, Bingham (1982) *Contemporary Democracies: Participation, Stability, and Violence* (Cambridge: Harvard University Press).

Przeworski, Adam and Teune, Henry (1970) *The Logic of Comparative Social Inquiry* (New York: Wiley-Interscience).

Putnam, Robert, Leonardi, Robert and Nanetti, Raffaella (1993). *Making Democracy Work: Civic Traditions in Modern Italy* (Princeton University Press).

Ragin, Charles (1987) *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies* (Berkeley: University of California Press).

Ragin, Charles and Becker, Howard (1992) *What is a Case? Exploring the Foundations of Social Inquiry* (Cambridge University Press).

Rihoux, Benoit (2006) 'Qualitative Comparative Analysis (QCA) and Related Systematic Comparative Methods', *International Sociology,* 21 (5), 679–706.

Seawright, Jason and Gerring, John (2008) 'Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options', *Political Research Quarterly,* 61 (2), 294–308.

Singer, David (1961) 'The Level-of-Analysis Problem in International Relations', *World Politics: A Quarterly Journal of International Relations,* 14 (1), 77–92.

Skocpol, Theda (1979) *States and Social Revolutions: A Comparative Analysis of France, Russia, and China* (Cambridge University Press).

Tarrow, Sidney (2010) 'The Strategy of Paired Comparison: Toward a Theory of Practice', *Comparative Political Studies,* 43 (2), 230–59.

Van der Heijden, Jeroen (2012) 'Voluntary Environmental Governance Arrangements', *Environmental Politics,* 21 (3), 486–509

Van Waarden, Frans (1995) 'Persistence of National Policy Styles, in Brigitte Unger and Frans Van Waarden (eds) *Convergence or Diversity? Internationalization and Economic Policy Response* (Aldershot: Avebury), 333–72.

VanWynsberghe, Rob and Khan, Samia (2008) 'Redefining Case Study', *International Journal of Qualitative Methods,* 6 (2), 80–5.

Venesson, Pascall (2008) 'Case Studies and Process Tracing: Theories and Practices', in Donatella Della Porta and Michael Keating (eds) *Approaches and Methodologies in the Social Sciences* (Cambridge University Press), 223–39.

Verba, Sidney and Nie, Norman (1972) *Participation in America: Political Democracy and Social Equality* (New York: Harper and Row).

Vogel, Steven (1996) *Freer Markets, More Rules: Regulatory Reform in Advanced Industrial Countries* (London: Cornell University Press).

Yin, Robert (2003). *Case Study Research. Design and Methods*, 3rd edn (Thousand Oaks: Sage).

Young, Oran (1999). *The Effectiveness of International Environmental Regimes* (Cambridge: MIT Press).

Part II
Comparative Designs

4
Case Studies and (Causal-) Process Tracing

Joachim Blatter and *Markus Haverland*

Introduction

Case-study research has been defined by Yin as an in-depth investigation of (contemporary) phenomena in a real-life context, particularly equipped to answer how and why questions (2009: pp. 8–18). Yin and other authors of case studies offer various analytical strategies for studying one of a few cases in depth, ranging from theoretically informed pattern matching (Yin, 2009) to strongly inductive approaches (Stake, 1995). This chapter deals with one specific approach: Causal-Process Tracing (CPT). This methodological approach is particularly well suited to answer 'why' and 'how' questions because it focuses on the causal conditions, configurations and mechanisms which make a specific outcome possible. It is outcome (Y)-centred, which means that the researcher is interested in the many and complex causes of a specific outcome and not so much in the effects of a specific cause (X). In other words, CPT is geared to answer questions like 'why did this (Y) happen?' Furthermore, its aim is to reveal the sequential and situational interplay between causal conditions and mechanisms in order to show in detail how these causal factors generate the outcome of interest.

One reason for the interest in process tracing is that it is part of what Hall (2003) famously called the attempt to 'align methodology to ontology'. He pointed to the fact that many theories and theoretical frameworks, like Historical Institutionalism (HI) and Game Theory, emphasize the relevance of 'timing', 'interaction effects' and 'contexts'. Furthermore, Ragin (2008) as well as Goertz and Mahoney (2012) have emphasized the role of 'configurational thinking' in theory and methodology. CPT is uniquely suited to take these aspects into account. In consequence, it has more to offer than just enhancing the internal validity of explanations

that are generated on the basis of co-variational methods (as implied by Gerring, 2007: pp. 173–84).

CPT is a within-case method of analysis that concentrates on the processes and/or mechanisms that link the causes and the effects within specific cases. In contrast, comparative case studies analyze patterns of variation in the scores of the dependent variables with the variation in the scores of the independent variables across cases (Blatter and Haverland, 2012: ch. 2). The two case-study approaches can be combined (Blatter and Haverland, 2012: ch. 5), but as a stand-alone approach, CPT has not only distinct research goals and its own techniques of data generation and data analysis, the logic of case selection and generalization also differ substantially from those that are connected to the methods that draw causal inferences on the basis of cross-case comparisons (Blatter and Haverland, 2012: ch. 3).

This chapter will first illustrate that a method that focuses on the temporal unfolding of causality is part of the ongoing trend to align methodology to ontology. We will then discuss different understandings of process tracing and defend our approach, which has affinities to critical realism as the epistemological base and to two main pillars, that is, configurational causation and temporality. Next, we will introduce three distinct types of causal process observations, which form the empirical bases for drawing causal inferences within a CPT approach. The features of the CPT approach will be illustrated by examples from different fields of public policy.

The case for process tracing: Aligning methodology to ontology

Why is it fruitful to go beyond cross-case comparisons and to focus on within-case observations? Tracing evidence of processes within cases was always part of comparative case studies in public policy and related fields. While cross-case comparative designs, such as the Most Similar Systems Design, and their methods of case selection and causal inference have received a lot of methodological thought (for example, Przeworski and Teune, 1970; Lijphart, 1971, 1975) there was not much methodological reflection about the analysis of processes within cases. In fact, while some ideas about historical processes have been sketched in George's approach of structured focused comparisons (George, 1979), it was, to our knowledge, only in 1985 that the term 'process tracing' was introduced in the methodological debate and briefly elaborated upon (George and McKeown, 1985).

Yet many comparative case studies drew their strength less from the way they compared cases but more from the detailed storylines that were presented for each case. Take the probably most often discussed work in the comparative social sciences: Skocpol's *State and Social Revolutions* (1979). Many scholars argue that it is less her cross-case *comparison* that makes her study compelling but more her analysis of revolutionary *processes* (Mahoney, 1999: p. 1157; Sewell, 2005: p. 97). That is why she focuses on the causal conjunctures and causal chains that have led to the revolutions in France, Russia and China (Mahoney, 1999: pp. 1164–8).

More generally, discussing why comparative case studies have been so influential in comparative politics, despite the widely acknowledged limitation of comparative case-study designs, Hall argues that many scholars in comparative politics have not relied heavily on the comparative method, 'instead the success of the field rests largely on the ways it has moved beyond that method' (2004: p. 2). Scholars moved 'toward the view that explanation consists in identifying the causal process that lie behind political outcomes' (Hall, 2004: p. 3).

In a related but more comprehensive piece, Hall lays out that a greater orientation on causal processes is justified given our assumptions about the nature of the social world and its underlying causal structure (ontology). According to Hall, the ontologies of comparative politics 'have outrun its methodologies' (Hall, 2003: p. 398) and the focus on processes is an attempt to 'align methodology with ontology'. Theories of comparative politics, and we would add comparative public policy, specify complex causal structures. These specifications are not only 'incompatible with the assumptions required by regression analysis' (Hall, 2003: p. 399), but also with the assumption of the comparative case-study methods informed by Mill's method of agreement and difference that assumes that only one or a few causal variables matter and that causal relationships are deterministic (Hall, 2003: pp. 379–80).

In particular, scholars who analyze (long-term) process of policy change have often very explicitly emphasized the role of timing and of causal mechanisms. Much of their thought has revolved around the notion of path dependency. Path dependency points towards the importance of time, timing and sequences. Early events in a process are more decisive than later events. Path dependency denotes positive feedback loops or increasing returns that keep the development upon a certain path even though alternatives might be preferable. Positive feedback loops are usually illustrated with the QWERTY keyboard, which has been widely adopted although it might not be the most optimal structure for a keyboard. In political science, it was probably Paul Pierson who has

most systematically theorized about the mechanisms underlying path dependent processes, applying mechanisms identified by economists to the nature of the political world. Regarding public policy, Pierson emphasizes adaptive expectations, learning effects and coordination effects:

> Institutions and policies may encourage individuals and organizations to invest in specialized skills, deepen relationships with other individuals and organizations, and develop particular political and social identities. These activities increase the attractiveness of existing institutional arrangements relative to hypothetical alternatives. As social actors make commitments based on existing institutions and policies, their cost of exit from established arrangements generally rises dramatically.
>
> (Pierson, 2000: p. 259)

An important example includes public pension schemes, where contributors have strong expectations regarding income maintenance after retirement and have adapted their financial planning accordingly. Mature pension schemes embody a tremendous implicit public commitment to scheme members. As a result, changes in public pension schemes are electorally very costly, and therefore, politicians have stayed on the path and have proposed rather incremental changes (see Myles and Pierson, 2000).

Another important concept within HI is the notion of critical junctures. A critical juncture is a combination of permissive conditions – conditions that ease the constraints and hence allow for change, and productive conditions – conditions that produce the new outcomes or range of outcomes after the permissive conditions disappear (Soifer, 2010). These conditions come together at a specific point in time and are individually necessary and jointly sufficient to produce change.

Path dependency and critical junctures are theoretical concepts from HI that highlight the relevance of timing in causal explanations. The former points to the insight that causal processes at a certain point in time can be influenced by earlier processes, and the latter highlights the importance of causal combinations at a specific point in time. In order to develop more general notions that are not connected with HI, we talk about 'causal chains' when we trace the sequential combination of causal factors over time and introduce the term 'causal conjunctions' when we want to indicate that two or more factors interact at a specific point in time for producing a specific outcome. These two concepts highlight the importance of timing in causal explanations, and they are based

on the logic of 'configurational thinking' (see below). In consequence, they are more useful than notoriously fuzzy terms like 'complexity' or 'context' that we find often in the literature on case-study method or process tracing. This leads us to the next point – the danger that the term 'process tracing' is becoming the next excuse or escape clause for methodologically unreflective research.

Process tracing and causal-process tracing: From a broad term to a specified methodological approach

Many studies in comparative public policy and related areas claim to involve process tracing. There is however a certain danger that 'process tracing' is used as a fuzzy catch-all term, employed in the same way as the label 'case study' has often been used – dropping a name without a full reflection on what it is that one is actually doing, what it implies with regard to drawing causal inferences and which steps are involved between the research question and the conclusion – a verbal shortcut rather than a full-fledged design. Until recently, the main problem was that a comprehensive treatment of the process-tracing approach was lacking; the problem nowadays is that there are quite diverse understandings and descriptions (George and Bennett, 2005; Gerring, 2007; Bennett, 2008, 2010; Falleti and Lynch, 2009; Collier, 2011; Beach and Pedersen, 2012; Blatter and Haverland, 2012; Rohlfing, 2012) of the process-tracing approach.

George and Bennett's broad description

George and Bennett's *Case Studies and Theory Development* (2005) has been the start of an ongoing methodological debate about CPT and (causal) mechanisms. They introduced process tracing as 'an operational procedure for attempting to identify and verify the observable within-case implications of causal mechanisms' and defined 'causal mechanisms as ultimately unobservable physical, social, or psychological processes through which agents with causal capacities operate but only in specific contexts or conditions, to transfer energy, information, or matter to other entities' (George and Bennett, 2005: pp. 137–8). Furthermore, they stressed that explanations via causal mechanisms have affinities to a scientific realist epistemology. This basically means that it is assumed that the reality exists independent of the observer; that it implies a commitment to micro-foundations – fundamental assumptions about the behaviour of actors, which form the basis for most basic social science theories; and that explanations via causal mechanisms draw on spatial contiguity and temporal succession (George and Bennett, 2005: pp. 137–45).

Towards a more specific understanding

The emerging methodological debate has been characterized by the fact that all scholars have emphasized some of these characteristic elements and downplayed others. It is not possible to delve very deep into the methodological debate here. In the following section, we will present an internally consistent and comprehensive approach that we call 'causal-process tracing' and make clear where we depart from alternative under-standings of 'process tracing'. In consequence, we use the term 'process tracing' as an unspecified umbrella term and 'causal-process tracing (CPT)' as a specified version. We are deep-hearted pluralists and want to stress the fact that, very often, the differences are a matter of emphasis and that the 'correct' approach depends very much on the specific research goals and the ontological and epistemological perspectives that dominate in a field of research. Therefore, the first basic message for students and scholars who want to apply CPT is: be explicit about your understanding of this approach and precise with respect to the method-ological source to which you refer to!

Configurational thinking

The CPT approach is based on configurational thinking (Ragin, 2008: pp. 109–46). This means that we connect the CPT strongly to the set-theoretical approaches and methods (Qualitative Comparative Analysis or QCA in its divergent variants) that have made strong inroads into the way we conceptualize cross-case comparisons in medium-N studies (Rihoux and Ragin, 2009). Configurational thinking implies that we start our analysis with the following assumptions: (i) social outcomes are the result of a combination of causal factors; (ii) there are divergent pathways to similar outcomes (equifinality); and (iii) the effects of the same causal factor can be different in different contexts and combinations (causal heterogeneity). In consequence, within a CPT approach, we talk about 'causal conditions' and not about 'variables' as in co-variational approaches. Furthermore, a CPT approach means that we search for the individually 'necessary' and jointly 'sufficient' conditions for the specific outcome in the case(s) under investigation. Finally, it implies that we do not generalize the findings to a population of similar cases, but we contribute to the debate on which causal pathways and configurations make a certain kind of outcome possible.

A configurational approach strikes a balance between the universalist way of thinking that characterizes co-variational methods (which assume that the causal power of 'independent variables' is autonomous and always and everywhere the same) and the particularistic way of thinking that

is implied with the term 'context'. In other words, it takes 'contingency' into account but forces the researcher to reflect on the specific status that 'context factors' have in the explanation of an outcome.

Causal mechanisms

Configurational thinking leads to a specific understanding of 'causal mechanisms' as a specific kind of causal configuration. We connect our understanding of causal mechanism to the macro-micro-macro-model of social explanation that is favoured by those scholars who want to link empirical research to basic social theory. In consequence, a causal mechanism is defined as a configurative entity that combines three different types of social mechanisms: 'situational mechanisms', 'action-formation mechanisms' and 'transformational mechanisms' (Esser, 1993, 1999–2001; Elster, 1998; Hedstroem and Swedberg, 1998: p. 22; Hedstroem and Ylikoski, 2010). Please note that this conceptualization of causal mechanisms does not imply a commitment to strong versions of methodological individualism, nor does it exclude structural factors as explanations as some critics assume (Hedstroem and Ylikoski, 2010: pp. 59–60).

Such a configurational understanding of causal mechanism differs from the understanding of those who assert that mechanisms are nothing but intervening variables (King et al., 1994); it is also distinct from a 'mechanismic' definition that is applied by Beach and Pedersen (2012). Some core elements of a mechanismic definition are fully consistent with our configurational understanding, for example, if a mechanism is defined as 'a set of interacting parts – an assembly of elements producing an effect not inherent in any of them' (Hernes, 1998: p. 78, according to Beach and Pedersen, 2012: p. 44). But neither the machine analogy that is used in order to illustrate such an understanding nor the overly general notion that 'each of the parts of a causal mechanism can be conceptualized as composed of entities that undertake activities' (Beach and Pedersen, 2012: p. 45, with further references) is in any way helpful for connecting the method of CPT to an accumulative process of social theory building as it is the case with the three kinds of social mechanisms that we propose to be the configurational elements of a causal mechanism.

Temporality

Our emphasis on temporality takes up George and Bennett's insight that spatial contiguity and temporal succession are important episte-mological foundations for drawing causal inferences (George and

Bennett, 2005: p. 140). These aspects have been elements of Hume's approach to prove causality, but they have been almost totally neglected in comparison to his third element: the regular association/co-variation among variables across cases. Unfortunately, in recent treatments of process tracing, Bennett (2008, 2010) and Collier (2011) do not highlight these 'naturalist/realist' foundations anymore – instead, they focus on tests for deductively derived hypotheses and on the logic of Bayesian updating when they describe process tracing. There is nothing wrong with a deductive, theory-guided approach to case studies and with introducing Bayesianism to case-study research; we have done the same under the heading of an approach that we call 'congruence analysis' (Blatter and Haverland, 2012: ch. 4). Nevertheless, if we take terminology seriously, (causal-*) process* tracing should be a method in which time and temporality play a major role (an emphasis that is shared by Rohlfing, 2012: p. 164). Of course, spatial contiguity and temporal continuity should not be applied in a naïve way but in combination with theoretical concepts, which have flourished in recent years (for example, Mahoney, 2000, 2006; Pierson, 2000, 2004; Büthe, 2002; Grzymala-Busse, 2011). This means, for example, that 'spatial contiguity' can also be translated into 'close ties' within a network-analytic theoretical framework. Only an understanding of CPT that accepts these 'realist' foundations as necessary (but not sufficient) conditions for drawing causal inferences can be a helpful complement to more deductive approaches to case-study research (Blatter, 2012). Finally, an approach that combines 'configurational thinking' with temporality is especially suited to complement QCA. It starts with similar ontological assumptions and adds a temporal dimension to this rather static method of cross-case analysis – only knowledge about timing makes a list of necessary and sufficient ingredients a full-fledged 'recipe'!

Causal-process observations as empirical bases for drawing inferences

We start this section by presenting two different definitions of causal-process observations. These definitions mirror nicely the differences between an unspecified understanding of process tracing and our CPT approach. This comparison represents a kind of summary of the last section and as a bridge to the following section where we present three kinds of causal-process observations. These observations form the empirical basis for drawing inferences about the status of causal factors and about the kind of dynamics that take place in a causal process.

Different definitions of causal-process observations

Causal-process observations have been defined in the following ways:

> An insight or piece of data that provides information about context, process, or mechanism, and that contributes distinctive leverage in causal inference. A causal-process observation sometimes resembles a smoking gun that confirms a causal inference in qualitative research, and is frequently viewed as an indispensable supplement to correlation-based inference in quantitative research as well.
>
> <div align="right">(Seawright and Collier, 2004: p. 283)</div>

> A cluster of empirical information that is used (a) to determine the temporal order in which causal factors work together to produce the outcome of interest; (b) to determine the status of these causal factors as individually necessary and jointly sufficient for the outcome in the cases under investigation; and/or (c) to identify and to specify the social mechanisms that form the basis for mechanism-based explanations.
>
> <div align="right">(Blatter and Haverland, 2012: p. 23)</div>

Three types of causal-process observations

In our book we have specified three types of causal-process observations, which are helpful to fulfil the three goals that are contained in our definition: comprehensive storylines, smoking-gun observations and confessions (Blatter and Haverland, 2012: pp. 110–19).

(i) A small-N study based on CPT should provide a 'comprehensive storyline' in which the development of potentially relevant causal conditions is presented in a narrative style. A major goal of these comprehensive storylines is to differentiate the major sequences of the overall process and identify the critical moments that further shape the process.

(ii) The study should also provide more detailed insight into the causal processes that occur at critical moments. The most important goal is to find empirical evidence that provides a high level of certainty that a causal factor or a combination of causal factors actually led to the next step in the causal pathway or to the final outcome of interest. In other words, we attempt to find smoking-gun observations embedded in a dense net of observations that show the temporal and spatial proximity of causes and effects. Note that, in contrast to authors such as van Evera and Bennett, we argue that a smoking-gun observation does not stand alone and is not used as a

test of a specific hypothesis (van Evera, 1997; Bennett, 2010). In our conceptualization, a smoking-gun observation is connected to other observations, and together, the full cluster of observations can be used inductively to make strong causal claims. Hence, a smoking-gun observation receives its strength for making causal inferences by its dense temporal and spatial connection to other empirical observations and not by its connection to a specific theory or hypothesis. A gun is an especially important piece of evidence if we observe it when it is still smoking following its use with a significant consequence (for example, killing somebody). In other words, the metaphor refers to temporal contiguity between the observation that points to a specific cause and other observations that provide evidence about the consequence(s) of that cause.

(iii) In order to gain 'deeper' insights into the perceptions, motivations and anticipations of major actors, we need further observations that we call 'confessions' because we want to highlight the complementary role of these 'observations' to the smoking-gun observations and we want to indicate that scholars applying a CPT approach should think like attorneys, who have to convince juries, and not so much like statisticians. Please note that confessions are important pieces of evidence, but as in judicial trials, we should not take them at face value without critical reflection. We should carefully examine the contexts in which actors provide information about their perceptions, motivations and anticipations. For example, when actors are interviewed by journalists or scholars, processes of ex-post rationalization often occur: actors justify their decisions by arguing that they pursued a specific goal, but in reality, the behaviour was much less reflective and strategically oriented, or it was driven by other goals. On the other hand, statements that actors make within the social or political processes often serve strategic purposes: they attempt to send signals to other actors to enhance their bargaining power or to strengthen their legitimacy to a wider audience. In other words, we should be aware of typical biases with respect to motivations when we interpret the statements of actors in specific contexts.

Confessions are important complements to smoking guns because they reduce the problem of drawing causal inferences on the basis of temporal succession. Actors can anticipate certain developments or actions and react to these anticipated developments in advance. This undermines the logic of drawing causal inference on the basis of temporal succession because the 'cause' lies ahead of the 'consequence'. Nevertheless, with respect to logic, the problem can easily be solved

because the 'real' sequence is as follows: (i) stimulus, which triggered the anticipation; (ii) action in accordance with the anticipation; and (iii) adjustment to or avoidance of the anticipated situation. The real challenge lies at the empirical level, especially when the anticipated situation does not occur because of earlier adjustments. Nevertheless, in principle, it is possible to identify the first 'critical moment' at which the actor began to change his behaviour in anticipation of a situation that he perceived to be possible or probable.

These three types of causal-process observations build the empirical basis for a thorough reflection on the question of whether certain causes or configurations should be viewed as necessary or sufficient causal conditions for the outcome in the case under investigation. Furthermore, they provide the necessary empirical information for drawing conclusions with respect to the kind of process dynamics that play an increasing role in recent theorizing. We will provide some examples in the following section, but first we have to make clear that empirical observations alone are never sufficient for drawing inferences. They have to be complemented by and interpreted with the help of further logical and theoretical tools.

Conclusions for the status of causal factors as necessary and sufficient conditions

How can we bolster the claim that a causal factor should be assigned the status of a necessary condition or the claim that a causal factor or a causal configuration has been sufficient for an outcome through within-case analysis? Different kinds of empirical observations and different techniques and theoretical concepts are necessary for these tasks.

The quest for necessary conditions is much more X-centric than the search for sufficient conditions. As Goertz and Levy (2007: p. 15) explain: 'To say that X is necessary for Y means simultaneously the counterfactual that without X, Y would not have occurred'. In consequence, if we want to make the claim that a factor has been a necessary condition, we have to apply counterfactual reasoning. Counterfactual thought experiments have been described mostly within a co-variational logic of causal inference (for example, Tetlock and Belkin, 1996; Blatter and Haverland, 2012: pp. 76–8), but they are not restricted to independent variables. Within a CPT approach, they can be applied to all major steps within a causal chain. Crucially important is the insight that counterfactual reasoning requires a solid and comprehensive knowledge of the historical process that we want to 'rewrite' in our thought experiment.

In contrast, if we want to bolster the claim that a causal factor (or a configuration of causal factors) has been sufficient for an outcome (or for the next step in a causal chain), we should turn towards coherent

theoretical models based on a consistent set of social mechanisms. These theory-based multilevel models of causation provide a logically consistent 'pathway' from a causal factor to an outcome. Furthermore, each element of these models – each social mechanism – is deterministic, but overall, the outcome is contingent on the specific configuration of social mechanisms. Therefore, we need empirical observations for each step to clarify which situational, action-formation and transformational mechanisms have been operating. Smoking-gun observations and confessions are the most important observations for these kinds of information. Together, the multilevel models of causation and the dense and deep insights that we achieve through smoking-gun observations and confessions provide the best foundation for making strong claims about sufficient conditions.

Process dynamics

Bennett and Elman (2006) have made us aware that positive feedback loops, which form the logical underpinnings of the concept of path-dependency, are not the only process dynamics that we should take into account when we try to explain social outcomes. Negative feedback loops and cyclical processes are also very common. Bennett and Elman (2006: p. 258) illustrate these alternative dynamics with the balance-of-power dynamics in the Westphalian State System and with the politics of abortion – each success of the proponents of abortion resulted in an increased mobilization of the opponents, and vice versa. These alternative process dynamics are driven by underlying causal mechanisms that can be aligned to basic social theories (see Mahoney, 2000).

We can specify the roles that the different kinds of causal-process observations play to put empirical flesh on the logical bones of these process dynamics. The comprehensive storylines are necessary to identify which kind of dynamics has actually occurred. This reveals how important it is, within a CPT approach, to justify the period of time that we take into account in our empirical study. It is possible that a process that exhibits strong features of path dependency, based on mechanisms that provide positive feedback loops in a shorter period of time, is much more accurately described as a cyclical process if we take a more long-term perspective.

Identifying the process dynamics with the help of comprehensive storylines is only the first step in a causal explanation that focuses on these dynamics. The next step is to trace the causal mechanisms that lead to positive and/or negative feedback loops. For this task, we rely on the kind of empirical information that smoking-gun observations and confessions represent.

Examples

In the following section, we provide a few examples to illustrate the elements of the CPT approach that we laid out before. They cover main areas of public-policy research – the welfare state and the regulatory state, but also environmental policy making. The first two examples are introduced to highlight the role of temporality for drawing causal inferences; the second example also covers smoking guns and confessions. The next two examples are included in order to show that CPT can concentrate on development at the macro level but that a CPT-based study can also focus on mechanisms, which means that it combines the macro level with the micro level. At the end, we come back to our first example and show how CPT can complement QCA studies.

Emmenegger (2010) – The long road to flexicurity: The development of job security regulations in Denmark and Sweden

In this comparative case study on job security regulations in Denmark and Sweden, Patrick Emmenegger scrutinized the historical development of job security regulations from the end of the nineteenth century to 1966 in the first section of his article. Then, in two following sections, he provided more detailed descriptions of what happened in the two countries after 1966. The temporal structuration of his analysis is justified because he refers to path dependency, critical junctures and windows of opportunity as major elements of his explanatory framework. Before 1966, the two countries were on the same path, which was characterized by a corporatist solution to job security since the beginning of the twentieth century in contrast to Continental Europe, where job security was early on regulated by law. He chooses to have a deeper look into the cases after 1966 because, at this time, both countries were governed by left-wing majorities and had the opportunity to depart from their earlier path. Whereas Sweden turned towards legal regulations, Denmark did not. Emmenegger explains these divergent developments by describing how the strategies of the labour movement developed over time in both countries and how the turn towards stronger demands for job security regulations in Sweden coincided with the dominance of the Social Democrats in Sweden, whereas the Social Democrats in Denmark had already lost their powerful position again when the same chance in labour strategy happened in Denmark. Here the window of opportunity for legal regulations was already closed when the labour unions changed their strategy and Denmark stayed on the established track.

The Emmenegger examples show the importance of timing when two factors are seen as necessary and jointly sufficient for producing an

outcome. The change in job security policy occurred only in that country in which the demand from the labour unions and the dominance of the Social Democrats in the parliament happened at the same time. In other words, the identified causal configuration is situational, and therefore, it can be specified as a causal conjunction. The next example shows that timing is also important when we want to make the argument that specific conditions worked together sequentially in a causal chain.

Blatter (2009) – Performing symbolic politics and international environmental regulations: Tracing and theorizing a causal mechanism beyond Regime Theory

In his single case study, Blatter combines CPT with more deductive case-study methods. First, he shows that all explanatory approaches associated with Regime Theory in International Relations (IR) as well as the Advocacy Coalition Framework (ACF), developed for analyzing domestic environmental policy making, fail to explain the regulations of motorboats at Lake Constance, a regulation that stimulated similar regulations by the European Union and Switzerland. The ACF explains why the motorboat regulations were always high on the policy agenda of political institutions that were created by the national and subnational governments of the riparian states of Germany, Austria and Switzerland; the transboundary space had seen the emergence of two broad and powerful but antagonistic coalitions, which fought for and against these regulations. Blatter argues that the regulatory breakthroughs at the beginning of the 1970s and the beginning of the 1990s cannot be explained by IR theories because neither favourable interest constellations nor other facilitating factors were given, and he argues that the policy-centric ACF is also not able to capture the real momentum that made far-reaching regulations possible. In order to bolster his argument that the regulations have to be understood as a result of political attempts to symbolize the emergence of a new polity (the transnational Euregio Bodensee), Blatter turns first to timing. He shows that the breakthroughs took place when the European-wide idea of Euro-region building washed over the shores of Lake Constance and not during those times when the issues of motorboat emission were high on the public agenda in the cross-border region. Next, he shows that the *Internationale Bodensee Konferenz* (IBK), an institution that was set up by the subnational governments of the neighbouring states for signalling their will to dominate the emerging transnational political space, played a major role in the negotiations that lead to the breakthroughs. This in contrast to the intergovernmental commissions, which were set up by the national

governments and which were officially in charge of the regulation of the international lake. The close connection between the IBK activities and the breakthroughs can be seen as a smoking-gun observation for his theoretical claim that transnational region building and not intergovernmental environmental policy making was the decisive trigger for the breakthroughs. This is complemented by statements that resemble 'confessions', for example, the claim in the IBK announcement of the regulations that this 'pioneering work indicates that regions are of special importance in Europe, they give important impulses and offer solutions' (translations from Blatter, 2009: p. 100). The environmental coalition was able to put the issue on the agenda in the 1960s, and again in the 1980s, but there it was blocked by the user coalition. When the subnational governments were looking for an issue they could use in order to show their relevance during the high times of Euro-region building, they took up the unsolved motorboat issue and pushed the regulations through. Overall, in order to understand how the pioneering regulation was possible, we have to look at the entire causal chain, whereby each factor was necessary to push the process further but only together were they sufficient for creating innovative cross-border regulations.

Hacker (1998) – The historical logic of national health insurance

Hacker's analysis of development of health systems in Britain, Canada and the United States serves as an example of a process-tracing study that operates at the macro level. In line with our emphasis on temporality and configurational thinking, Hacker stresses that health policy outcomes are not the result of a few autonomous independent variables and that the timing – in particular the timing of government policies – and the sequence of events are of crucial importance. He states that

> policy feedback, state building, the timing and sequence of institutional change – these are not 'values' of a particular variable but rather complex causal processes that unfold historically and involve large numbers of interlocking, often inseparable causes.
>
> (Hacker, 1998: p. 84)

The study first describes the structural causal conditions that may have an influence of the development of health policies – medical complexity, societal interests and the structure of the political systems. The discussion of these conditions serves, on the one hand, to debunk straightforward mono-causal explanations of health policy development, and, on the other hand, as an ingredient for Hacker's more complex explanation.

For this more complex explanation, he uses the concepts of 'path depen-
dence' and 'critical junctures'. Critical junctures are produced for
instance by partisan shifts or significant external shocks. Yet these critical
junctures are – in our words – only necessary and not sufficient condi-
tions for policy change. Whether policy change does indeed occur
depends on specific interactions between 'the incentives and constraints
created by political institutions, and the inherited legacies of past policies'
(Hacker, 1998: p. 127).

Hacker then constructs comprehensive storylines for each of his three
countries. These storylines span more than a hundred years, and it
comes as no surprise that his article has the unusual length of 74 pages.
For each country, he demonstrates how prior policies interact with other
(changing) conditions to produce new outcomes. The story-lines are
structured around critical junctures. In the United Kingdom, for
instance, such a critical juncture was the landslide victory of the Labour
party in 1945. This victory happened against the end of World War II
that had increased state capacity vis-à-vis society and nurtured social
solidarity. This conjunction of factors provided a fertile ground for putting
the idea voiced in the 1942 Beveridge Report into practice by creating
the National Health Service. In line with his theoretical argument, the
detailed and focused discussion of this critical juncture is preceded by an
analysis of the prior policies, in particular, the 1911 National Insurance
Act and its legacy.

Hacker identifies policy decisions (or non-decisions) regarding three
issues that have shaped political struggles in the three countries he had
studied. It is worthwhile to present them here because they demonstrate
that these kinds of findings cannot be established with comparative
public-policy designs – neither large-N nor small-N (comparative case
studies).

The first policy decision was whether or not *private* health insurance
was allowed to develop and the form it took. If the schemes developed
were 'physician-friendly' in which many – middle-class – citizens were
enrolled, then physicians would subsequently strongly oppose government-
sponsored schemes as would the middle-class citizen who had become
attached to the initial schemes. Moreover, these schemes would result in
relatively high costs of medical care while at the same time it would be
more difficult to enact national health insurance (Hacker, 1998: p. 82).

Second, it matters at which target group *public* insurance programmes
were initially aimed. For instance, in the United Kingdom, public insurance
programmes were set up for the working class, while in the United States,

it solely focused on poor citizens, the disabled and the elderly (Medicaid and Medicare). Hence, in the United States, the focus was on those groups that are difficult and costly to insure. Therefore, a lot of resources needed to be committed to those groups making it more difficult to expand programmes to other groups. Also, it freed 'the private sector to focus on those populations for which private insurance is most viable' (Hacker, 1998: p. 82).

The third choice concerns 'the relative timing of public efforts to bolster the technological sophistication of medicine, on the one hand, and to increase access of citizens to health care, on the other' (Hacker, 1998: p. 83). In the United States, the government has put a lot of effort into improving health technology. This has made health provision quite costly. Subsequently, the issue of increasing access to health care has resulted in a particularly strong tension between costs and access, which has served as an impediment to universal health insurance (Hacker, 1998: p. 83).

Overall, Hacker's study illustrates a number of important points about CPT. First of all, as already mentioned, the study makes clear that public policy outcomes cannot be fully explained by just focusing on independent and dependent variables as large-N studies and the comparative case-study methods do. Only through within-case analysis, can causal processes and feedback mechanisms be identified. The study also shows how comprehensive storylines can be constructed, alternating between less intensive discussions of longer time periods and more intensive discussions of critical junctures. At the same time, the study remains at the macro level. It does not delve much into the mechanisms that work at the individual level, or at least, the author does not provide evidence that the mechanisms have worked as hypothesized.

Kuran and Sunstein (1999) – Availability cascades and risk regulation

In contrast to the previous example, Kuran and Sunstein focus on mechanisms linking the macro level with the individual micro level. Theirs is a study of public mobilization and risk regulation. It deals with mass scares about comparatively minor risks and how they translate into legislation and into the commitment of considerable resources. As stated above, this study contrasts nicely with the one by Hacker. Not only is this a study in the realm of the regulatory state rather than the welfare state, it also focuses on the micro-level mechanisms, rather than the macro-level sequences as Hacker's study does.

The mechanisms utilized in this study are cognitive and social mechanisms that shape how individuals develop beliefs about certain issues that, in turn, shape their risk perception and preferences. More specifically, the study develops an argument about a number of interacting mechanisms that are mediated by what is called the availability heuristic, an idea from social psychology that the 'perceived likelihood of any given event is tied to the ease with which the occurrence can be brought to mind' (Kuran and Sunstein, 1999: p. 685).

There are two mechanisms involved in the study and both are concerned with how individuals form judgements and preferences. For both mechanisms, the authors assume that choices of individuals are dependent on the choices of other individuals. Moreover, under certain conditions, these choices become self-reinforcing, a positive feedback loop or, in the words of the authors, a 'cascade' (Kuran and Sunstein, 1999).

Informational cascades occur 'when people with incomplete personal information on a particular matter base their beliefs on the apparent beliefs of others'; hence, people may accept a belief 'simply by the virtue of its acceptance by others' (Kuran and Sunstein, 1999: p. 686). In addition to the information cascade, they introduce the notion of the reputational cascade. They argue that individuals are not only influenced by others because others are more knowledgeable but because they seek to gain social approval and avoid social disapproval (Kuran and Sunstein, 1999: p. 686). Accordingly, they communicate and behave as if they share what they think is the dominant belief (Kuran and Sunstein, 1999: p. 687). According to the authors, both cascades do not develop in isolation from each other. Rather they interact or reinforce each other typically triggered by a salient event, resulting in a so-called 'availability cascade' (687).

The last elements of their theory are 'availability entrepreneurs'. Availability entrepreneurs are actors anywhere in the political system who understand these dynamics and seek to trigger availability cascades to advance their own cause (Kuran and Sunstein, 1999: p. 687). They seek to set these cascades into motion by directing the attention of others, including the public, to specific problems by interpreting phenomena in specific ways and by seeking to increase the salience of certain information (Kuran and Sunstein, 1999: p. 687).

Kuran and Sunstein try to trace the working of the mechanisms by a number of case studies. Their most detailed study concerns 'Love Canal'. Love Canal is a canal in New York that was filled with chemical waste by a company in the 1940s and early 1950s and then covered with dirt. Later, the area around the canal was used by the local government to build 200 houses. The site of the old canal itself became a

school and a playground. Twenty years later, a governmental commission found chemicals in fishes in Lake Ontario and identified Love Canal as the major source. This served as a triggering event for an availability cascade.

In a comprehensive storyline, the study describes the different stages through which the availability cascade develops. Inhabitants of the area around Love Canal started to get frightened about possible health effects. Although there has never been sound scientific evidence that there are health effects, the availability cascade unfolded when one of the inhabitants in that area acted as an availability entrepreneur. The inhabitant directed attention to the issue. She organized a petition drive and went door-to-door to mobilize protest. She was also critical in debunking scientific evidence. The inhabitants of the area met frequently, formally and informally, and accepted her belief that the situation was dangerous. This provides evidence for the claim that risk perceptions are formed dependently. The whole local community became close-knit, indicating that the reputational mechanism would 'convince' those who might still have doubts. The worries of the community were taken up by the New York State Health Commissioner, the Governor and later even President Jimmy Carter. Their actions, aimed at precaution, were interpreted by the citizens as proof of the seriousness of the problem, and scientific reports to the contrary were criticized and mistrusted. The availability entrepreneur was very successful in increasing the salience of the issue. She appeared on national television and was invited to Congress and the White House. The availability cascade, as such, was not the only cause for far-reaching regulation, but it tipped the balance in favour of adopting the Comprehensive Environmental Response, Compensation and Liability Act (CERCLA) or the Superfund statute which called for 1.6 billion in expenditure to clean potentially hazardous abandoned waste sites with many billion more to follow in subsequent years (Kuran and Sunstein, 1999: pp. 691–8).

In order to support the claim of the working of causal mechanisms at the individual level, the study also includes confessions. Some confessions take the shape of quotes taken from two books written by the availability entrepreneur herself. For instance, she revealed how she sought to pressure a public official into action:

> I jumped up and said to Commission Whalen: 'If the dump will hurt pregnant women and children under two, what for God's sake, is it going to do to the rest of us!? ... Now very emotional, I said: 'You can't do that! That would be murder!'
>
> (Kuran and Sunstein, 1999: p. 693)

These quotes provide for an in-depth insight of her perceptions and motivations and, hence, constitute important traces of the causal mechanisms at work.

Combining CPT with QCA

Finally, we want to come back to our first example in order to show that an understanding of process tracing, which combines configurational thinking with an emphasis on temporality, makes a perfect within-case complement to medium-N studies, which combine configurational thinking with set-theoretical methods of cross-case comparisons (QCA). Emmenegger has embedded the case studies on job security regulations in Denmark and Sweden (Emmenegger, 2010) within a medium-N study applying QCA (Emmenegger, 2011b). He conducted his case studies as a follow-up to the QCA analysis and used the results of the QCA analysis for selecting the cases for the in-depth study. The outcome in Denmark represents a contradictory or non-consistent case with respect to the findings of the QCA analysis (Emmenegger, 2011b). According to the QCA analysis, which included 19 Western democracies, the existence of high levels of non-market coordination, strong labour movements, moderately strong religious parties and few institutional veto points represent one of three identified causal configurations that led to a high level of job security regulations. Denmark displayed all these conditions but did not follow the other Nordic countries towards a high level of job security regulation. By applying CPT, Emmenegger is able to show that timing made the difference. Although Denmark had all the ingredients that were sufficient for a high level of job security regulation in other countries, they did not come together at the right time.

Emmenegger selected not only the non-consistent case, Denmark, but also a consistent one, Sweden, for his timing-centred case study. Thus, he followed the advice of Schneider and Rohlfing (2013), who recommend a comparative process tracing study after a QCA study. Using Sweden as a comparative case that exhibits the same causal configuration as Denmark and high levels of job security regulations (hence, a consistent case for the Nordic path) helps in various ways to focus the tracing of the causal process in Denmark. First, the comparison makes it possible to identify the time when the regulatory paths of the two countries diverged, and second, the Swedish case provides a template of necessary and jointly sufficient factors for a successful turn towards higher regulations. Emmenegger (2011a: p. 19) contains a figure that nicely visualizes the causal chain of necessary and sufficient conditions in the

Swedish case. With the help of this template, Emmenegger is able to identify the small but crucial differences between the two countries with respect to the strength of leftist parties and with respect to time when the unions changed their strategy.

Two insights should be stressed when we reflect on how causal inferences are drawn in Emmenegger's comparative case study. The cross-case comparison helps not only to exclude many theoretically possible explanations for the outcome in Denmark, but also is helpful in identifying causal chains and causal conjunctions, which represent temporal configurations of necessary and jointly sufficient causal conditions. Nevertheless, the basis for drawing causal inferences is located in the within-case analysis, and the cogency of Emmenegger's conclusion depends to a large part on a dense description of the temporal unfolding of events and on information and reflections about the perceptions and anticipations of core actors.

Finally, the examples reveal that CPT is an especially valuable methodological tool when we want to solve 'puzzles' (Grofman, 2001) or investigate 'theoretical anomalies' (Rogowski, 2004) or search for 'omitted variables' or 'scope conditions' by looking at cases that are 'deviant' or 'outliers' in the light of a large-N statistical study or that are 'contradictory cases' according to the findings of a medium-N QCA (Gerring, 2007: pp. 105–15; Blatter and Haverland, 2012: p. 211). By definition, the claim that a specific factor has been necessary and (together with other conditions) sufficient to produce the unexpected, deviant or contradictory outcome cannot be based on the comparison of evidence with expectations that have been deduced from prior knowledge or theories. In consequence, we want to emphasize our conviction that an understanding of CPT that has strong affinities to 'scientific realism' and that draws heavily on the fact that causality plays out in time and space makes this method an especially valuable inductive complement to more deductively oriented research methods.

Conclusion

Many theoretical developments in public policy analysis explicitly or implicitly highlight the importance of timing and temporality. Causal-process tracing is the most important move in small-N studies to re-align methodology to such an ontology. Unfortunately, recent attempts to clarify the method of process tracing have downplayed its scientific realist foundation and the important role that spatio-temporal proximity plays

in providing (inductive) leverage for an explanation. In this chapter, we have laid out an understanding of CPT that is based on two pillars: configurational thinking and timing. We introduced the notions of causal conjunctions and causal chains as time-focussed specifications of causal configurations that help us to overcome broader and therefore more fuzzy notions like context, contingency or complexity; and we argued for an understanding of causal mechanisms as a configuration of three kinds of social mechanisms because such an understanding will facilitate the accumulation of knowledge and theory building in the social sciences. Furthermore, we showed that comprehensive storylines, smoking-gun observations and confessions represent different kinds of causal-process observations that provide the necessary empirical information for drawing causal inferences with respect to the necessary and sufficient conditions for the outcome of interest in research designs that are based on within-case analysis. Such an understanding of CPT allows for complementing large-N studies (based on statistical methods) and medium-N studies (based on QCA). Our specific understanding makes it also a helpful tool for complementing more deductive-oriented small-N methods, like co-variational analysis – an approach that draws causal inferences on the basis of co-variations among variables across cases; and congruence analysis – an approach that draws inferences by comparing the congruence between a large set of empirical observations and diverse expectations that we can deduce from a plurality of theoretical frameworks (Blatter and Haverland, 2012).

Bibliography

Beach, Derek and Rasmus Brun Pedersen (2012) *Process Tracing Methods – Foundations and Guidelines* (University of Michigan Press).

Bennett, Andrew (2008) 'Process Tracing: A Bayesian Perspective', in Janet M. Box-Steffensmeier, Henry E. Brady and Collier David (eds) *The Oxford Handbook of Political Methodology* (Oxford University Press), 702–21.

Bennett, Andrew (2010) 'Process Tracing and Causal Inference', in Henry E. Brady and David Collier (eds) *Rethinking Social Inquiry*, 2nd edn (Lanham: Rowman and Littlefield), 207–19.

Bennett, Andrew and Colin Elman (2006) 'Complex Causal Relations and Case Study Methods: The Example of Path Dependence', *Political Analysis*, 14 (3), 250–67.

Blatter, Joachim (2009) 'Performing Symbolic Politics and International Environmental Regulation: Tracing and Theorizing a Causal Mechanism beyond Regime Theory', *Global Environmental Politics*, 9 (4), 81–110.

Blatter, Joachim (2012) 'Taking Terminology and Timing Seriously: On Ontological and Epistemological Foundations of Causal-Process Tracing', Paper presented at the ECPR Joint Sessions, 10–14 April 2012, Antwerp. [http://www.unilu.ch/files/blatter-taking-terminology-and-timing-seriously-ecpr-antwerp.pdf]

Blatter, Joachim and Markus Haverland (2012) *Designing Case Studies. Explanatory Approaches in Small-N Research* (Houndmills: Palgrave Macmillan).

Büthe, Tim (2002) 'Taking Temporality Seriously: Modeling History and the Use of Narrative and Evidence', *American Political Science Review*, 96 (3), 481–93.

Collier, David (2011) 'Understanding Process Tracing', *Political Science and Politics*, 44 (4), 823–30.

Elster, Jon (1998) 'A Plea for Mechanisms', in Peter Hedstroem and Richard Swedberg (eds) *Social Mechanisms: An Analytical Approach to Social Theory* (Cambridge University Press), 45–73.

Emmenegger, Patrick (2010) 'The Long Road to Flexicurity: The Development of Job Security Regulations in Denmark and Sweden', *Scandinavian Political Studies*, 33 (3), 271–94.

Emmenegger, Patrick (2011a) 'How Good Are Your Counterfactuals? Assessing Quantitative Macro-Comparative Welfare State Research with Qualitative Criteria', *Journal of European Social Policy*, 21 (4), 365–80.

Emmenegger, Patrick (2011b) 'Job Security Regulations in Western Democracies: A Fuzzy Set Analysis', *European Journal of Political Research*, 50 (3), 336–64.

Esser, Hartmut (1993) *Soziologie: Allgemeine Grundlagen* (Frankfurt am Main: Campus).

Esser, Hartmut (1999–2001) *Soziologie: Spezielle Grundlagen*, vol. 1–6 (Frankfurt am Main: Campus).

Falleti, Tulia G. and Julia F. Lynch (2009) 'Context and Causal Mechanisms in Political Analysis', *Comparative Political Studies*, 42 (9), 1143–66.

George, Alexander L. (1979) 'Case Studies and Theory Development: The Method of Structured, Focused Comparison', in Paul Gordon Lauren (ed.) *Diplomacy: New Approaches in History, Theory and Policy* (New York: Free Press), 43–68.

George, Alexander L. and Andrew Bennett (2005) *Case Studies and Theory Development in the Social Sciences* (Cambridge: MIT Press).

George, Alexander L. and Timothy J. McKeown (1985) 'Case Studies and Theories of Organizational Decision Making', *Advances in Information Processing in Organizations*, 2 (1), 21–58.

Gerring, John (2007) *Case Study Research: Principles and Practices* (Cambridge University Press).

Goertz, Gary and Jack S. Levy (2007) 'Causal Explanation, Necessary Conditions, and Case Studies', in Gary Goertz and Jack S. Levy (eds) *Explaining War and Peace: Case Studies and Necessary Condition Counterfactuals* (London: Routledge), 9–46.

Goertz, Gary and James Mahoney (2012) *A Tale of Two Cultures: Contrasting Qualitative and Quantitative Paradigms* (Princeton University Press).

Grofman, Bernard (ed.) (2001) *Political Science as Puzzle Solving* (Ann Arbor: The University of Michigan Press).

Grzymala-Busse, Anna (2011) 'Time Will Tell? Temporality and the Analysis of Causal Mechanisms and Processes', *Comparative Political Studies*, 44 (9), 1267–97.

Hacker, Jacob (1998) 'The Historical Logic of National Health Insurance: Structure and Sequence in the Development of British, Canadian and U.S. Medical Policy', *Studies in American Political Development*, 12 (1), 57–130.

Hall, Peter A. (2003) 'Aligning Ontology and Methodology in Comparative Politics', in James Mahoney and Dietrich Rueschemeyer (eds) *Comparative Historical Analysis in the Social Sciences* (Cambridge University Press), 373–404.

Hall, Peter A. (2004) 'Letter from the President. Beyond the Comparative Method', *Newsletter of the Organized Section in Comparative Politics of the American Political Science Association*, 15 (2), 1–4.

Hedstroem, Peter and Petri Ylikoski (2010) 'Causal Mechanisms in the Social Sciences', *Annual Review of Sociology*, 36 (August 2010), 49–67.

Hedstroem, Peter and Richard Swedberg (1998) 'Social Mechanisms: An Introductory Essay', in Peter Hedstroem and Richard Swedberg (eds) *Social Mechanisms: An Analytical Approach to Social Theory* (Cambridge University Press), 1–31.

King, Gary, Robert O. Keohane and Sidney Verba (1994) *Designing Social Inquiry: Scientific Inference in Qualitative Research* (Princeton University Press).

Kuran, Timur and Cass R. Sunstein (1999) 'Availability Cascades and Risk Regulation', *Stanford Law Review*, 51 (4), 683–768.

Lijphart, Arend (1971) 'Comparative Politics and the Comparative Method', *American Political Science Review*, 65 (3), 682–93.

Lijphart, Arend (1975) 'The Comparable-Cases Strategy in Comparative Research', *Comparative Political Studies*, 8 (2), 158–77.

Mahoney, James (1999) 'Nominal, Ordinal, and Narrative Appraisal in Macrocausal Analysis', *The American Journal of Sociology*, 104 (4), 1154–96.

Mahoney, James (2000) 'Path Dependence in Historical Sociology', *Theory and Society*, 29 (4), 507–48.

Mahoney, James (2006) 'Analyzing Path Dependence: Lessons from the Social Sciences', in A. Wimmer and R. Kössler (eds) *Understanding Change: Models, Methodologies, and Metaphors* (Houndmills: Palgrave Macmillan), 129–39.

Myles, John and Paul Pierson (2000) 'The Comparative Political Economy of Pension Reform', in P. Pierson (ed.) *The New Politics of the Welfare State* (New York: Oxford University Press), 305–33.

Pierson, Paul (2000) 'Increasing Returns, Path Dependence, and the Study of Politics', *American Political Science Review*, 94 (2), 251–67.

Pierson, Paul (2004) *Politics in Time: History, Institutions, and Social Analysis* (Princeton University Press).

Przeworski, Adam and Henry Teune (1970) *The Logic of Comparative Social Inquiry* (New York: Wiley and Sons).

Ragin, Charles C. (2008) *Redesigning Social Inquiry: Fuzzy Sets and Beyond* (University of Chicago Press).

Rihoux, Benoît and Charles C. Ragin (eds) (2009) *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques* (Thousand Oaks: Sage).

Rogowski, Ralf (2004) 'How Inference in the Social (but not the Physical) Sciences Neglects Theoretical Anomaly', in Henry E. Brady and David Collier (eds) *Rethinking Social Inquiry* (Lanham: Rowman and Littlefield), 75–83.

Rohlfing, Ingo (2012) *Case Studies and Causal Inference. An Integrative Framework* (Houndsmills: Palgrave Macmillan).

Schneider, Carsten Q. and Ingo Rohlfing (2013) 'Combining QCA and Process Tracing in Set-Theoretic Multi-Method Research', *Sociological Research and Research*, 42 (4), 559–97.

Seawright, Jason and David Collier (2004) 'Glossary of Selected Terms', in Henry E. Brady and David Collier (eds) *Rethinking Social Inquiry: Diverse Tools, Shared Standards* (Lanham: Rowman and Littlefield), 273–313.

Sewell, William H. (2005) *The Logics of History: Social Theory and Social Transformation* (Chicago University Press).

Skocpol, Theda (1979) *States and Social Revolutions: A Comparative Analysis of France, Russia, and China* (Cambridge University Press).

Soifer, Hillel David (2010) 'The Causal Logic of Critical Junctures', *Political Methodology Committee on Concepts and Methods Working Paper Series*, vol. 24 (January 2010).

Stake, Robert E. (1995) *The Art of Case Study Research* (London: Sage).

Tetlock, Philip and Aaron Belkin (eds) (1996) *Counterfactual Thought Experiments in World Politics: Logical, Methodological, and Psychological Perspectives* (Princeton University Press).

van Evera, Stephen (1997) *Guide to the Methods for Students of Political Science* (Ithaca: Cornell University Press).

Yin, Robert (2009 [1984]) *Case Study Research* (Thousand Oaks: Sage).

5
Intermediate-N Comparison: Configurational Comparative Methods

Isabelle Engeli, Benoît Rihoux and *Christine Rothmayr Allison*

Introduction

Configurational Comparative Methods (CCMs), also called 'set-theoretic' methods, is a broad and encompassing label that embraces crisp-set and fuzzy-set Qualitative Comparative Analysis (QCA) as well as some alternative techniques. All CCM-related techniques conceive cases as configurations of attributes and are geared towards systematic cross-case analysis. QCA represents both a distinctive research approach, with its own aims and set of assumptions, and an umbrella term for specific techniques such as *fuzzy-set QCA* (*fsQCA*), which will be the focus of this chapter (Ragin, 2000, 2008a; Rihoux and Ragin, 2009; Schneider and Wagemann, 2012; Thiem and Dusa, 2013). In Ragin's seminal book (1987), the CCM approach was launched by the development of a new technique in which investigation was based on grouping dichotomous cases, crisp-set QCA (csQCA). It was later developed into fsQCA and other related techniques. The comparative essence of QCA stems from the fact that it was initially geared towards the analysis of multiple cases in a small- and intermediate-N research design (Marx et al., 2013). Ragin's motivation (1987, 1997) was to develop a '*synthetic strategy*' as a middle way between the case-oriented (or 'qualitative') and the variable-oriented (or 'quantitative') approaches. According to Ragin, this middle way would '*integrate the best features of the case-oriented approach with the best features of the variable-oriented approach*' (Ragin, 1987: p. 84). On the one hand, QCA aims at contextualizing explanations by gathering in-depth within-case knowledge and capturing the complexity of cases. On the other hand, if the ability of QCA to approximate 'the best features' of regression analysis approaches is contested (see, for instance, Lieberson, 2004; Seawright, 2005; Hug, 2012), QCA nevertheless provides some systematic tools for comparing

an often limited number of policies that are available for investigation. QCA concentrates on producing some level of parsimony across cases and, thereby, facilitates limited forms of middle-range or 'contingent' generalization for medium-N analysis in comparative public policy (Ragin, 1987; Blatter and Blume, 2008; Blatter and Haverland, 2012).[1]

On the qualitative side, QCA embodies some key strengths of the case-oriented approach (Ragin, 1987; Berg-Schlosser et al., 2009). To start with, it is a holistic or contextualizing approach and each individual case is considered as a complex *configuration* of causal conditions – the explanatory factors – and an *outcome* – the phenomenon to be explained. In addition, QCA is based on a conception of causality that leaves room for complexity (Ragin, 1987; Berg-Schlosser et al., 2009). QCA scholars often emphasize the idea of *multiple conjunctural causation* to best describe the ontological foundation of the approach. Thinking in terms of multiple conjunctural causation implies three key features. First, instead of singling out the explanatory power of individual factors, multiple conjunctural causation assumes that it is often a combination of causal conditions that eventually produces the outcome. Second, it posits that different combinations of conditions may lead to a similar outcome. Third, the impact of an individual condition may differ according to how it is combined with the presence or absence of other conditions. In other words, QCA does not assume any permanent causal direction between a given condition and the outcome. Drawing on Mill's major piece *A System of Logic* (1967 [1843]), Ragin pleads in favour of moving away from mono-causality and argues instead that causality is context- and conjuncture-sensitive. While using QCA, the researcher is steered away from developing a single causal model that best fits the data and urged to '*determine the number and character of the different causal models that exist among comparable cases*' (Ragin, 1987).

On the quantitative side, QCA should be better thought of as a 'complement' to regression analysis – and a 'remedy' for a medium-N analysis – rather than an approximation of or a 'competitor' to regression analysis (Mello, 2013: pp. 1–2). QCA has been specifically designed to handle the analysis of a medium-N in a systematic way, an issue that has been somewhat neglected in the past. It offers a middle way between the comparative study of a small number of cases and the quantitative treatment of a large number of cases (Moses and Knutsen, 2007). Comparative public policy scholars often face challenges regarding the sample of observations to include in the analysis: the N is either too small to run robust statistical analysis or too large to conduct in-depth qualitative studies on a whole set of countries or regions. For instance, in examining

the determinants of gender equality policy outputs in Western Europe, any analysis is immediately constrained by the number of empirical cases available. One option is to select a few representative cases according to some of the qualitative case-selection strategies developed in Chapter 3 in this volume. However, gender equality policies usually consist of a complex set of various regulations addressing several policy fields such as reproductive rights, discrimination in the labour market and gender roles in the family. As such, selecting only a few cases necessarily obliges the researcher to omit a portion of the variation in the dependent variable. The other commonly applied research design strategy is to include all Western European countries in the analysis. One would then face some methodological robustness and validity issues in running common statistical analyses, such as regression analysis, due to the small-N. Relying on QCA-related techniques allows for increasing the range of cases included in the analysis while aiming at some middle-range generalization. Through some minimization algorithms, QCA allows for a smoother and more systematic identification of regularities. As such, in contrast with classic comparative case studies, it opens up the possibility of achieving more parsimonious explanations for qualitative comparative research on a larger number of cases. In addition, QCA relies on formalized analytical tools and thus requires some forms of modelling that can be replicated (Berg-Schlosser et al., 2009). This replicability enables further verification of the results of the analysis and so places QCA into a more cumulative knowledge research approach.

QCA as a relevant methodological approach for comparative public policy

CCMs, and here QCA-related techniques, can be used to achieve two different purposes (De Meur and Rihoux, 2002; Berg-Schlosser et al., 2009). The most basic use is simply to summarize data, to describe cases in a synthetic way, and to use QCA as a tool for data exploration and data description. In the same manner, it can also be used to check coherence within the data: the detection of contradictions (the so-called 'contradictory configurations' in QCA terminology) allows one to learn more about the individual cases, in a dialogue between QCA and case-based knowledge (Rihoux and Lobe, 2009). For instance, McVeigh et al. (2006) examine the ways through which social movements and non-profit organizations can affect the agenda-setting process on issues related to hate crimes in the United States. They use csQCA to classify and map out the different channels of influence and the conditions under which

these organizations can successfully impact agenda setting. The second and more ambitious purpose of QCA is to test competing existing theories and hypotheses and show how different explanations could actually turn out to be more complementary than competing. As QCA relies on the assumption of multiple conjunctural causation, it provides some systematic tools for theory-testing in qualitative research. Drawing on fsQCA, McBride and Mazur (2010) have shown that the drivers for promoting women's rights are multiple and context-sensitive. Failing to identify any single factor that could exert a net and systematic effect, they instead reveal how different factors – drawn from competing explanations found in the literature – may actually intertwine and better explain when and why state-based promotion of gender equality has been successful over time. In this perspective, QCA can provide, to a certain extent, support for theory refinement. However, QCA is ill-equipped for analytic induction. If QCA enables discovery through a dialogue with the data and allows researchers to build upon case-based knowledge, it is not an inductive tool dedicated to generate theoretical propositions (Hug, 2012). In short, deep theoretical knowledge should drive each step of QCA.

From the late 1990s onwards, an increasing number of comparative public policy analyses have started to resort to QCA-related techniques.[2] This choice is based on the need to gather in-depth insight into the different cases and capture their complexity while still attempting to produce some level of limited generalization (Ragin, 1987). The preoccupation with within-case complexity is well in line with the growing trend towards 'evidence-based' policy analysis that, in principle, requires not only standard numerical (quantitative) indicators across the units of observation or 'cases', but also qualitative and 'thick' knowledge embedded in the actual cases (Yin, 2008). The preoccupation with generalization is obviously directly compatible with what a lot of public policy analysis is about: performing diagnoses across many observations, and deriving generalizations – or at least some systemization in the analysis – that could lead to the formulation of policy prescriptions (for example, the appropriate policy instruments or 'policy mixes' to achieve a given policy goal).

This trend towards multiple case studies also coincides, more generally, with a renewed interest in rigorous forms of case-oriented research (Mahoney and Rueschemeyer, 2003; George and Bennett, 2005; Gerring, 2007), and with new attempts to engage in a more productive dialogue between the 'quantitative' and 'qualitative' empirical traditions (Sprinz and Nahmias-Wolinsky, 2004; Moses et al., 2005; Brady and Collier,

2010). Such a strategy is particularly well-adapted for policy-related research since many potential objects of study – from the viewpoint of both scholars and policy practitioners – are 'naturally' limited in number: nation-states or regions, different kinds of policies in different states, policy outputs and outcomes, policy programmes, policy styles, policy sectors. These naturally limited or 'small-N' (or 'intermediate-N') populations are often especially relevant from a policy perspective (Rihoux and Grimm, 2006). The overall strength of QCA is it offers some systematic tools for going beyond loose *post hoc* comparison of the case study material typically gathered in the field of public policy analysis.

QCA-related techniques offer practical added value for comparative public policy analysis in any sector in at least three different ways (Rihoux and Grimm, 2006). First, these techniques allow for systematically comparing policy programmes in a small- or intermediate-N design, with across countries, regions or sectors. These may be typically within or across broad political entities or groups of countries (for example, the European Union, the OECD, a given category of countries for a regional comparison), but also within-country comparisons (for example, across states in the United States, across *Länder* in Germany) or within-region comparisons at more disaggregated levels (for example, between economic basins, municipalities). Second, these techniques allow for testing ex-post and ex-ante alternative explanatory (policy intervention) models leading to a favourable/unfavourable policy output and favourable/unfavourable policy outcomes (Varone et al., 2006). This approach allows the identification of more than one unique path to a policy outcome; more than one combination of conditions may account for a unique result. This is extremely useful in real-life policy practice, as experience shows that policy effectiveness is often dependent upon national/regional settings as well as upon sector-specific features, and that different cultural, political and administrative traditions often call for differentiated implementation schemes (Audretsch et al., 2005). Third, these techniques enable the policy analyst or policy evaluator to examine under which conditions (or more precisely, under which combinations of conditions) a specific policy would be effective or not.

QCA as a set of techniques

In this section, we address some key features and issues related to the application of configurational comparative methods to the analysis of comparative public policy. To illustrate the discussion, we rely on an fsQCA of the regulatory regimes on human biotechnology conducted

by Engeli and Rothmayr (2013). Through this empirical illustration, we discuss some key stages in conducting fsQCA in comparative public policy – along broad lines, similar stages would need to be followed for a csQCA or a multi-value QCA (mvQCA).

Why opt for CCMs to study regulatory regimes on human biotechnology?

Engeli and Rothmayr's study (2013) investigates the spatial and temporal variation in the regulatory regimes for human biotechnology (embryonic stem cell research) in 15 Western European countries. Since the early stages in the development of human biotechnology, regulatory variations across Western Europe have constituted a puzzle for comparative public policy scholars. Human biotechnology presents some promising avenues for developing treatments for life-threatening diseases such as Huntington's disease or boosting organ transplant possibilities. Given the economic and scientific potential of human biotechnology and increased international competition in R&D activities, purely economic and interest-driven accounts of policy development over time would predict minimal state regulation combined with strong promotional activities. Contrary to this expectation, though, the regulation of human biotechnology is highly divergent across Western Europe. Some countries such as Belgium or the United Kingdom have adopted minimal state regulation and strongly promoted research and development. In contrast, other countries such as France, Germany or Italy have opted for highly restrictive regulations that severely limit or even fully ban any research on embryonic stem cells.

The choice of CCMs – and in particular fsQCA – was driven by two main motivations. The first motivation was related to the intermediate number of available observations. The regulation of human biotechnology constitutes a typical case of medium-N analysis that is frequently encountered in comparative public policy research. On one hand, the number of Western European countries is invariably small; in addition, human biotechnology is a relatively recent field that only took off at the end of the 1970s and, consequently, the number of regulations is too small to fulfil the minimal requirements for running any robust large-N statistical analysis. On the other hand, the range of variation in the regulatory regimes is nevertheless relatively broad and several different types of regulatory regimes have been put in place to address human biotechnology issues. A classic comparative case-study analysis on a few countries would not have taken into account this regulatory diversity. The second motivation was that previous research on human biotechnology

policies differed greatly on the determinants of the regulatory variations and singled out alternative competing explanations for these regulatory variations that have not been tested in a simultaneous way. These contradictory results reveal at best an incomplete account of why and how policies diverge over time. Above all, they point at the need to move beyond monocausality and to investigate, instead, to what extent different configurations of conditions may actually produce a similar outcome. CCMs (and here fsQCA) represent a viable alternative to qualitative case studies for successfully dealing with medium-N. FsQCA allows for a better assessment of the multicausal nature of the regulatory regimes for human biotechnology by revealing how different causal processes may produce similar results without requiring a large number of observations that do not exist in the real world.

Crisp set versus fuzzy set? How to select configurational comparative techniques

Configurational comparative analysis covers a family of techniques. Two of them are increasingly used in comparative policy analysis: csQCA that was the initial technique designed by Ragin (1987) and fsQCA that was later developed to overcome some weaknesses of csQCA.[3] The two techniques share an overarching principle that is key to understanding how any QCA-related technique works: they all follow a set-theoretic approach (Ragin, 2008a; Schneider and Wagemann, 2012). In other words, QCA techniques investigate case membership in sets (Schneider and Wagemann, 2012). A case or an empirical observation is considered as being a member – or not a member – of an overarching qualitative set. The researcher has to carefully decide what the set is according to her research question. For instance, if one investigates democratic regimes, the United Kingdom would be considered as part of the 'parliamentary regime' set, while the United States – being a presidential system – would be considered as not being part of that set.[4] Here, the set is dichotomous in nature as a regime is or is not part of the 'parliamentary regime' set. This type of a dichotomous set is called *crisp set* in QCA terminology: cases are *either* members of the set *or* not members of the set. Crisp-set QCA, more commonly called csQCA, is the appropriate technique to deal with crisp sets. An increasing number of studies have successfully used csQCA (for a review of public policy analysis drawing on csQCA, see Rihoux et al., 2011).

While the advantages of csQCA are not contestable, it is nevertheless sometimes difficult to maintain a strict dichotomization while defining cases. Considering the parliamentary regime example, the classification of

France, Finland or Portugal would be challenging. Their semi-presidential system shares characteristics of both parliamentary and presidential regimes. One solution would be to leave France out of the analysis. Another solution is to move from crisp sets to fuzzy sets. A fuzzy set allows for partial membership in a set, and here France, Finland or Portugal would be considered as partial members of the 'parliamentary regime' set but not full members (as the United Kingdom would be). Fuzzy sets allow for more fine-tuned qualification of the cases, which is often called for in comparative public policy analysis. Take the classic example of evaluating the impact of a particular policy. As policy analysts well know, a policy never eventually achieves all its goals. It sometimes only fulfils part of its goals, it sometimes produces unexpected but positive effects, or it may also deploy some negative effects. Forcing policy impacts into a dichotomous categorization of 'impactful' versus 'non impactful' policies would distort the complexity of policy impact evaluation. On the contrary, adopting a fuzzy-set perspective on case membership allows for defining different degrees of policy impact and presenting a more fine-grained picture of the complexity in policy implementation.

A similar motivation drives Engeli and Rothmayr's (2013) choice of running fsQCA to analyze regulatory regimes for human biotechnology. While qualitatively investigating the various regulatory regimes set up across Western Europe, they came to the conclusion that the variety in regulatory regimes could not be captured by a simple dichotomy featuring 'permissive' regulations versus 'restrictive' regulations. Indeed, they found several regimes that incorporate some of both permissive and restrictive features. For instance, Switzerland has a complete ban on any human embryo research, whereas the United Kingdom even permits the creation of human embryos for research purposes. Switzerland and the United Kingdom can thus be easily classified as being fully out of the 'permissive regulation' set for the former, fully in the 'permissive regulation' set for the latter. However, some other countries have adopted a more intermediate approach to governing human biotechnology and fall in between the ideal types of Switzerland and the United Kingdom. The Netherlands, for instance, bans embryo creation for research purposes but nevertheless allows a large spectrum of research on existing embryos. If it is obvious that Dutch regulation is very different from the Swiss one, it is nevertheless not completely similar to the British one. Indeed, a regulation that authorizes a large amount of research on existing embryos may be permissive, but it still does not equate with a fully permissive regulation that would allow for embryo creation. The British and Dutch permissive regulations are thus *qualitatively* different from

the restrictive Swiss regulation, while the Dutch and British regulations are also *quantitatively* different from each other in their degree of permissiveness. As a result, Engeli and Rothmayr (2013) opted for running fsQCA analysis for two main reasons. First, the fuzzy set approach allows them to qualitatively distinguish between various levels of permissiveness and not only between two opposite qualitative stages (permissive vs. restrictive). Second, they collected data that were precise enough to support a fine-grained distinction in the degree of permissiveness of the regulation, namely *restrictiveness*. They developed an index for evaluating the degree in permissiveness of the different regimes that was composed of three different dimensions. The three dimensions were first additively aggregated and then calibrated into a QCA fuzzy set, as we will explain in the following section.

In sum, choosing between crisp set versus fuzzy is a matter of the particular research question to be addressed as well as the quality and precision of the data available. CsQCA and fsQCA are less different in nature than in their capacity to apprehend complexity in the cases. If a simple dichotomy is sustainable for a particular research question, csQCA is the right technique to use. If the observed variation cannot be apprehended into just two categories – member of the set versus non-member of the set – and if there is precise enough data to adopt a more fine-grained categorization, fsQCA becomes the appropriate choice.

Tips and tricks in calibrating: Operationalization in fsQCA analysis

Once the decision over crisp sets versus fuzzy sets has been made, the next crucial stage is the operationalization of the equivalent of 'dependent' and 'independent' variables in quantitative analysis. In QCA terminology, the dependent variable is called the *outcome*, while the potential explanatory factors are called the *conditions* – note that the latter are not 'independent' variables as they are expected to operate in combination. The process of operationalizing the conditions and outcomes into QCA membership sets is referred to as the *calibration* process (Ragin, 2008a, 2008b). In other words, calibration implies determining the extent to which a case belongs or does not belong to a theoretically defined set of cases and then assigning the corresponding membership scores. There are four types of calibration: (i) calibration of a dichotomous variable into a crisp set; (ii) qualitative or ratio or interval scale variable into a crisp set; (iii) qualitative variable into a fuzzy set; and (iv) ratio or interval scale variable into a fuzzy set. While the outcome variable has to take the form of a fuzzy set for running fsQCA analysis, the condition variables do not necessarily all have to be fuzzy sets. A mix of crisp and fuzzy sets in the conditions is

perfectly possible and is often a pragmatic way to deal with some data unavailability and measurement quality issues. An important preliminary decision to be made before calibrating addresses the number of conditions to include in the analysis. QCA was initially developed to handle small- and medium-N, and there is no fixed ratio between the number of conditions and the number of cases. Some benchmarks have, however, been established with regard to the *maximum* number of conditions that can be included, in connection with the number of cases, in order to avoid ill-specified csQCA models (Marx, 2010; Marx and Dusa, 2011).[5] Reviewing published QCA applications in different fields, Rihoux et al. (2013) find that most studies rely on an intermediate number of cases (between 10 and 50 cases) and the usual number of conditions ranges between four and six.

Whatever the format of the conditions, all QCA sets share some key common characteristics in fixing set-membership values. First, membership set scores range from the starting point of 0 to the end point of 1, which constitute *qualitative anchors* in QCA terminology. These two qualitative states represent the full membership of a particular set of interests (value 1), and the full non-membership of the same set (value 0). Second, along with the starting and ending anchors, there is another important qualitative anchor: 0.5, the *indifference value* or, in other words, the cross-over point. Setting a case at the cross-over point means that this particular case is neither more a member nor more a non-member of the set. All the cases assigned to a value higher than 0.5 are considered as being at least partial members of the respective set, while all the cases assigned to a value lower than 0.5 are considered, at best, partial non-members of the same set. If assigning the 0.5 value to any empirical case should be avoided in running the analysis, it is nevertheless important to theoretically define what the cross-over point is while calibrating the fuzzy set. The cross-over point determines the anchor above and below which any case is – relative to the other cases included in the analysis – more or less of a member of the set than another case. Third, a fuzzy set can scale any value greater than 0 but smaller than 1, which will represent the relative variation in degree in membership. In a set-theoretic approach such as fsQCA, a case is not included in a membership set by its own characteristics, but is *relatively included* in the set in relation to the characteristics of the other cases included in the set. For instance, assigning the value 0.4 to a case means that this particular case is rather not a member of the set in comparison to some other cases *but is not fully out of the set either*. In other words, the case presents some

characteristics of the set but not enough to be considered as being a member of the set. In the same vein, a value of 0.8 would mean that the case tends towards being a member of the set while still not possessing enough characteristics to be considered a full member (and thus be assigned the value of 1). Labels have been developed to facilitate the description of the degree of membership in a fuzzy set (Table 5.1).

Determining the number of values of a fuzzy set may be challenging. Theoretically, a fuzzy set could take an infinite number of values between 0 and 1 as long as the different values represent a substantive and theoretically meaningful difference in the degree of membership of a set. What constitutes a set, what the starting and end points mean, and what the cross-over point represents have to be defined at the theoretical level in QCA. There is no 'ready-to-apply' formula that could work for any and every set regardless of the theoretical justification for the calibration of the particular membership set. However, most of the time comparative public policy research faces some data unavailability and measurement quality issues that prevent the use of a large series of values. Empirically, most research tends to calibrate fuzzy sets to four (for instance, 0–0.3–0.7–1) or to six values (for instance, 0–0.2–0.4–0.6–0.8–1), with 0.5 always remaining the cross-over point. A case assigned with a value that is below 0.5 is partially or entirely not a member of the set, while a case assigned with a value above 0.5 is partially or fully a member of the set.

Again, Engeli and Rothmayr's (2013) study can be used to illustrate the fsQCA calibration process. In their investigation of the determinants of human biotechnology regulatory permissiveness, Engeli and Rothmayr analyze, among other possible causal factors, the impact of left-wing parties. The Social Democrats have been important members of cabinets across Western Europe over the last two decades, sometimes in coalition

Table 5.1 Labelling the degree of membership in a 6-value fuzzy set

Fuzzy Set Value	Description of the membership status
1	Full membership of the set
0.8	Partial membership: mostly but not fully in
0.6	Partial membership: more in than out
0.4	Partial non-membership: more out than in
0.3	Partial non-membership: mostly but not fully out
0	Full non-membership of the set

Source: Ragin (2008b)

governments. As precise and accurate data measuring the strength of left-wing parties in government is available, Engeli and Rothmayr were able to calibrate a fuzzy membership set. They opted for a four-value set in order to qualitatively distinguish to what extent Social Democrats were in a strong governmental position at the time of the regulation adoption, relying on Armingeon et al. (2011)'s indicator capturing the total percentage of cabinet positions held by Social Democratic and other leftist parties. The original indicator ran from 0 per cent (none) to 100 per cent (all) of cabinet positions. More importantly, the original variable contained four important gaps – between 0 per cent and 22.25 per cent, between 32.16 per cent and 50 per cent, 55.56 per cent and 78.95 per cent, as well as between 79.18 per cent and 100 per cent. Calibrating the original interval-scale variable implies a decision about what variation is to be captured by the set. One possible choice would have been to opt for a direct translation of the original interval-level scale into a fuzzy set that would have ranged a high number of values. If this direct translation would have yielded a very fine-grained measure, the theoretical meaning of these four important quantitative gaps described above would have been lost and it would have resulted in a potentially meaningless high number of values. The full range of quantitative variation would have been captured, but the relevant variation for the particular hypothesis of the study would have been disregarded. For these reasons, they chose instead to distinguish four qualitative types of governmental strengths and classify the cases according to the *degree of institutional leverage the Social Democrats could rely on to impose their policy preferences*. Drawing on the literature, it can be inferred that Social Democrats would exert weaker impact if they were a minor coalition member than if they were a major coalition member or fully in power. Accordingly, Engeli and Rothmayr set the cross-over point (0.5) between being a minor coalition member and major coalition member while the end-point value (1) was assigned to Social Democratic single-party government and the starting point (0) to governments with no Social Democratic members (Table 5.2).

Calibrating membership scores in QCA represents a necessary step towards a methodologically robust and theoretically valid analysis. The precise format of the membership set is not as important as the careful and transparent justification for the calibration itself. In the same way that regression analysis displays descriptive statistics to discuss the variables entered into the regression model, the calibration stage in QCA is a crucial prior stage to run a robust analysis.[6]

Table 5.2   Calibration of the governmental strength of Social Democrats

Cabinet positions %	Fuzzy-set score
0	0 'not in power' (fully out of the set)
22.25 to 32.16	0.33 'minor coalition member' (more out than in the set)
50 to 79.18	0.67 'major coalition member' (more in than out of the set)
100	1 'in full power' (fully in the set)

Source: Armingeon, et al. (2011)

Running an fsQCA analysis

While the basic analytical procedure of csQCA could, in principle, be computed by hand, as is the case for any basic statistics over a small number of observations, the analysis becomes increasingly complex as the number of cases and conditions increase. To facilitate the analysis and make it fully replicable, three main software packages have been designed to run csQCA and fsQCA: 'fs/QCA' 2.5 (Ragin and Davey, 2012) operating on Windows, 'QCA' 1.0–4 working in R (Thiem and Dusa, 2013) and 'Fuzzy' st0140_2 (Longest and Vaisey, 2008) for Stata.[7] The three packages compute the same general procedure with some differing features regarding graphics and advanced statistical routines. The study under discussion relies on the fs/QCA software, so far the most utilized of the three programmes. The description of the procedure is based on Ragin (2008a).

Constructing the truth table and minimizing the configurations of conditions

FsQCA analysis proceeds in two steps. The first step consists in constructing the so-called *truth table*. In technical terms, a truth table is a multidimensional vector space encompassing all the combinations among the different fuzzy sets modelled in the analysis. More simply, the truth table maps out the logically possible combinations of causal conditions that lead to the occurrence or non-occurrence of the outcome. To discuss the key features of a truth table, we rely on the following adapted version of Engeli and Rothmayr's truth table (2013).

Engeli and Rothmayr's study investigates the impact of six causal conditions (the 'configuration of conditions' columns going from 'religious parties' to 'ART') on the design of permissive regulation for human

biotechnology (the outcome). The full truth table (Table 5.3) displays all the logically possible combinations of conditions that would lead to a positive outcome – here, permissive regulation. The logical component of a QCA truth table should be stressed. A QCA truth table lists all the combinations of conditions that would be theoretically possible to produce the desired outcome. In other words, the truth table contains both configurations of conditions that empirically exist and configuration of conditions that are not represented by any empirical case but could still logically exist. As Engeli and Rothmayr's analysis covers six conditions, there are 64 logically possible combinations of conditions.[9] However, not all these combinations are supported by empirical evidence. The truth table classifies the membership scores distribution of their 22 empirical cases across the whole set of the 64 logical configurations of conditions. The 22 empirical cases included in their analysis cover only 14 of these logically possible combinations (rows 1 to 14). As column 'N' shows, some configurations of conditions are empirically supported by several cases (see, for instance, rows 1 and 2), while others correspond to just one case (see, for instance, row 3). As we can see the truth table is already a first important analytical step: it groups the cases that follow an identical combination of characteristics and thus produces a first synthesis. The truth table also shows the logical combinations that are logically possible but not supported by any of the empirical cases included in the analysis. In this example, there are 50 such combinations.

Table 5.3 FsQCA truth table[8]

Row	Causal conditions						Outcome consistency			N
	Religious Parties	LEFT GVT	Abortion	Science	Church	ART				
1	1	0	1	0	0	1	1	1	2	
2	0	1	1	1	0	1	1	1	3	
3	0	1	1	0	0	1	1	1	1	
4–6							1	1	…	
7	0	0	1	1	1	1	0	0.88	1	
8	0	0	0	1	0	0	0	0.52	2	
9–13							0	…	…	
14	1	1	0	1	0	0	0	0	1	
…							?	?	0	
64							?	?	0	

Frequency cut-off: 1.00; Consistency cut-off: 0.88

   The second step of the fsQCA consists of analyzing the truth table and reducing the logical complexity by the use of the Quine–McClusky algorithm (Ragin, 2008a) in order to identify necessary and/or sufficient causal path(s) leading to the outcome. This second step is called the *minimization* process in QCA terminology: the stepwise reduction of the long list of possible combinations of conditions that are contained in the truth table to the shortest possible expression – the *minimal solution*.

Making sense of the results: How to interpret causal paths

FsQCA computes three types of minimal solutions. The *parsimonious solution* is computed on the basis of both the empirical and logical cases (the cases that are logically possible but are not empirically represented in the set of cases included in the analysis). The *intermediate solution* keeps all the empirically observed cases and retains only *some* of the logical cases that are selected by the researcher in a theory-informed way.[10] The last solution is called the *complex solution*. Its minimization retains only the empirical cases and excludes all the logical cases. As such, the complex solution often provides less straightforward results to interpret. As the relatively small number of cases limits the minimization process, it often results in several and rather long causal paths in comparison to the parsimonious solution. Nevertheless, the complex solution is a 'safe option': it only relies on the empirical observations and does not imply the use of any unrealistic assumption that cannot be sustained at the theoretical level.

   Drawing on Engeli and Rothmayr's analysis of permissive regulatory regimes for human biotechnology (here considered as the outcome), Table 5.4 illustrates the differences between the parsimonious and the

Table 5.4 Alternative causal paths leading to the outcome (adapted from Engeli and Rothmayr, 2013)

Alternative paths	Consistency	Unique coverage
Parsimonious solution		
ART	0.81	0.93
Complex solution		
church*science*ART*LEFT GVT+	1.00	0.28
church*LEFT GVT*religious parties* ART +	1.00	0.35
church*ABORTION*left gvt*RELIGIOUS PARTIES* ART	1.00	0.14
CHURCH*ABORTION*SCIENCE*religious parties* ART	0.87	0.38
Solution coverage: 0.70; Solution consistency: 0.93		

complex solutions.[11] The parsimonious solution was computed on the 64 possible logical combinations of conditions, that is, the 14 configurations that were empirically sustained by at least one empirical case and the remaining (unobserved) 50 logical combinations. This parsimonious solution is extremely simple and straightforward to understand: the presence of only one single factor is necessary to produce the outcome (here the condition called 'ART') – this factor is, in fact, both necessary and sufficient for the occurrence of the outcome. On the contrary, the complex solution is more complicated: there are four alternate paths that are sufficient but not necessary to produce the outcome.[12] In addition, none of these causal paths are 'short' (parsimonious) as they all contain a conjunction of at least four conditions. This level of complexity frequently occurs when a fairly large number of conditions are modelled for a relatively small number of cases.

There is no rule that is set in stone for choosing one solution over the other. It is recommended to examine all three solutions (parsimonious, intermediate and complex) and to concentrate on the solution in which one has the most methodological confidence. Here, Engeli and Rothmayr decided to rely on the complex solution for two reasons. On the one hand, their N (=22) remained relatively small for running an fsQCA. However, it represents the totality of all existing human biotechnology regulations in Western Europe. It thus reduces the need for keeping the high number of logical but non-observed cases leading to the parsimonious solution. They thus considered the complex solution as being more robust than the parsimonious one, even if it was a more complicated one. On the other hand, their choice was reinforced by the fact that the complex solution presented no contradiction to the parsimonious solution. Indeed, in each of the four alternative paths of the complex solution, one finds the same condition that was singled out in the parsimonious solution (ART), combined with some other conditions. In other words, the complex solution is a subset of the parsimonious solution, and the parsimonious solution is a superset of the complex solution (Ragin and Sonnett, 2004). Besides these two stated reasons, a third reason that legitimizes their choice is that the complex solution, although still quite 'long', does produce *some* level of parsimony – especially the first two terms of the solution, and also, to a lesser extent, the third and fourth terms (see Table 5.4). In other words, each term of the QCA solution enables one to exclude some initial conditions from the explanation of the outcome, thereby moving beyond description.

Once the decision over which solution to focus on is made, it is then time to interpret the causal paths. As we have seen, Engeli and Rothmayr's

analysis reveals four alternative paths producing the outcome (the respective *terms* of the solution). This means that there is not necessarily any combination of conditions that would lead to the outcome but a series of alternative combinations, each of which being sufficient to produce the outcome '1'.[13] This series of alternative causal paths is very much in line with their theoretical expectation that posits that governing human biotechnology is a complex process triggered by different sets of factors across Europe. They argue that focusing on a single explanation is misleading and emphasize how factors combined in different ways may eventually result in similar policy outcomes. Indeed, the presence of four alternative paths emphasizes their theoretical argument and stresses one of the key features in CCMs: the assumption of *equifinality* (according to which different configuration of conditions may eventually produce a similar outcome). All the different paths are alternative sets of causal factors that deserve a similar level of attention from the researcher.

Causal paths are read as follows in fsQCA software. Capital letters mean the condition is present (for instance, 'LEFT GVT') in the causal path while small letters mean that the condition is absent ('left gvt'). The '*' indicates what is called a 'logical AND' in QCA terminology and represents the conjunction of two (or more) conditions to form a causal path leading to the outcome. As it is not the purpose of this chapter to discuss the substance of the results presented in Engeli and Rothmayr (2013), let us just spell out how to read the first causal path displayed in Table 5.4. In substance, the first causal path indicates that in secular countries ('church') the presence of a left-wing government ('LEFT') combined with the existence of permissive policies in related biotechnology fields ('ART') are likely to result in a permissive regulation over human biotechnology even if public opinion is not majoritarian in favour of supporting science advance in the field ('science').

If the substantive meaning of the causal paths is strongly based on theoretical expectation in any QCA analysis, there are nevertheless some goodness-of-fit measures that help to assess – to a certain extent – the robustness of the analysis.[14] At least two measures are important to take into account while evaluating the various causal paths (Ragin, 2006, 2008a): consistency and coverage. Consistency measures the extent to which a similar configuration of conditions leads to a similar outcome. In other words, the consistency measure assesses the degree to which an identical combination of causal factors may actually result in different outcomes – the outcome (all the fuzzy set scores of the outcome running above 0.5) and the negation of this outcome (all the scores ranging below 0.5). We use here the consistency measure developed by Ragin

(2003, 2006).[15] A perfect consistency score takes the value of 1.00, and it is generally assumed that consistency scores below 0.75 indicate some non-robust and unreliable configurational paths (Ragin, 2000, 2006). In the analysis under consideration here, the consistency score of the overall complex solution, shown in Table 5.4, is fairly high (0.93), and the weakest path's consistency score is 0.87. Overall, the level of consistency of the solution is, therefore, satisfying and lends confidence to the robustness of the results. The second coefficient widely used to assess the quality of the QCA solution is the coverage measure. This indicates the explanatory range of a causal path. In other words, it measures the number – or rather the proportion – of cases that display the same configurational path and provides some sense of the empirical relevance of the explanation. Generally speaking, the greater the proportion of cases that are covered by a particular causal path, the closer to 1.00 is the coverage measure of that path. Coverage scores can often be smaller than their respective consistency scores when there are several alternative paths leading to the outcome (for instance, the third path in Table 5.4). It is up to the researcher to decide to what extent these frequency considerations tapped by the coverage coefficient are important to consider in the interpretation. On the one hand, in small- or intermediate-N comparison, the general argument is that 'each causal path matters', regardless of frequency. In other words, even a causal path corresponding to one single 'special' case makes as much sense as another causal path with many cases – as QCA is in essence a non-probabilistic approach (Rihoux and Ragin, 2004). On the other hand, however, in the field of policy analysis, it could make sense to take frequency considerations into account, especially in the perspective of formulating policy recommendations for instance.

Conclusion: Intermediate-N analysis and beyond

This chapter introduced the core features of QCA analysis and how QCA can be successfully used in comparative public policy research relying on small and intermediate numbers of observations. All CCM-related techniques conceive cases as configurations of attributes and constitute a 'middle way' between the comparative study of a small number of cases and the quantitative treatment of a large number of cases. This middle way, which is often challenged by the number of observations that can and should be included in the analysis, fits some of the needs of comparative public policy research nicely. In comparative public policy research, the N is often too small to run robust large-scale statistical

analysis on its own and too large to conduct single-handed in-depth qualitative studies on the whole set of observations. Relying on QCA-related techniques allows for increasing the range of cases included in the analysis while aiming at some (limited) middle-range generalization.

There is an increasing trend towards using QCA in combination with other methods, either qualitative or quantitative (for a review, see Rihoux et al., 2011; Mello, 2013). Recent contributions discuss the advantages and limitations of consolidating large-N QCA and regression analysis (Amenta and Poulsen, 1996; Grofman and Schneider, 2009; Vis, 2012; Fiss et al., 2013). Another contemporary line of research looks at sequencing QCA and small-N qualitative analysis – such as process tracing (Rohlfing and Schneider, 2013) – to concentrate the qualitative investigation on a selection of cases that are identified as typical or deviant. Finally, QCA is moving out of its traditional association with cross-county analysis and is being employed to examine both meso-level (such as collective organizations) and micro-level (such as policy players) phenomena, once again demonstrating the versatility and utility of the technique for comparative policy research.

QCA is currently maturing into a methodological approach with its own properties for medium-N analysis in comparative public policy, while coexisting with other tools and techniques – qualitative case studies or regression analysis. The configurational/set logic that underpins QCA is unbounded by either the number of cases included in the analysis or the level of comparison. While some important weaknesses remain, QCA has gained its place as a distinct but complementary research strategy; equally so, comparative policy research has matured and prospered in the function of methodological rigor and the analytical purchase afforded by QCA.

Notes

1 We are careful not to use the term 'causal inference' here as its use has been contested for set-theoretic methods (Seawright, 2005).
2 For a recent review of the use of QCA in public policy analysis, see Rihoux et al. (2011), Rihoux et al. (2013), Emmenegger et al. (2013) and Mello (2013).
3 There is a third technique, mvQCA, which allows for multivalue causal conditions while the outcome has to remain dichotomous (Cronqvist and Berg-Schlosser, 2009).
4 This illustration on political system is based on Schneider and Wagemann (2012) and Ragin (2008a).
5 See also Marx (2010) and Marx and Dusa (2011) for an in-depth discussion of the validity issues raised by an unbalanced ratio between the number of

conditions and the number of cases. In particular, Marx and Dusa (2011) have developed benchmark tables for checking the probability of ill-specified csQCA models.

6   Ragin (2008a) has developed some statistical functions for calibrating quantitative variables into fuzzy set membership scores. For a discussion, see Ragin (2008a, 2008b), Schneider and Wagemann (2012) and Thiem (2010).

7   In addition, several other software packages are available. For a complete presentation and download information for the available software, consult: http://www.compasss.org/software.htm.

8   Rows 4–6 and 9–13: Due to length constraints, we only present here a truncated version of the truth table. The original truth table can be found in Engeli and Rothmayr (2013).

9   A vector space has 2k corners where k is the number of causal conditions. In Engeli and Rothmayr's analysis, the vector space has 64 corners (2^6), that is, 64 logical combinations.

10 Selecting the plausible logical remainders is a key – and often problematic – operation in constructing the intermediate solution that can involve theory-driven and case-driven considerations. For a discussion of the different available strategies to deal with limited diversity-related issues, see Schneider and Wagemann (2012), Ragin and Sonnett (2004), and Rihoux and Lobe (2009).

11 In the set-theoretic approach, patterns of causality are assumed to be asymmetric (no assumption of causal symmetry), and hence the respective configurations of conditions leading to the outcome and the negation of the outcome (where the negated outcome is ~O) can differ. Thus, it is strongly recommended to run the analysis for the negation of the outcome as well. For reasons of space, the results of the analysis for the negative outcome are not discussed in detail in this chapter.

12 A causal path is considered as *sufficient* if all the cases sharing the same causal path experience the same outcome. A causal path is said to be *necessary* if the causal path must be present for the case to experience the outcome.

13 This also means that each condition in those combinations of conditions is a so-called 'INUS' condition, that is, an *I*nsufficient but *N*on-redundant part of a combination of conditions which is itself *U*nnecessary but *S*ufficient to produce the outcome (Schneider and Wagemann, 2012).

14 For a discussion of the critical issues related to measurement error, misclassification and variable omission in deterministic causal analysis, see Lieberson (2004), Hug (2012), Braumoeller and Goertz (2000), Dion (1998), Maggetti and Levi-Faur (2013) and Seawright (2005).

15 Consistency $(X_i < Y_i) = \Sigma$ (min $(X_i, Y_i))/\Sigma$ (X_i), 'where "min" indicates the selection of the lower of the two values, X_i represents membership scores in a combination of conditions and Y_i represents membership scores in the outcome' (Ragin, 2006).

Bibliography

Amenta, Edwin and Jane D. Poulsen (1996) 'Social Politics in Context: The Institutional Political Theory and Social Spending at the End of the New Deal', *Social Forces*, 75, 33–61.

Armingeon, Klaus, David Weisstanner, Sarah Engler, Panajotis Potolidis and Marlène Gerber (2011) *Comparative Political Data Set I 1960–2009* (Institute of Political Science, University of Berne).

Audretsch, David B., Heike Grimm and Charles Wessner (eds) (2005) *Local Heroes in the Global Village: Globalization and New Entrepreneurship Policies* (New York: Springer).

Benoît Rihoux and Charles C. Ragin (2004) 'Qualitative Comparative Analysis (QCA): State of the Art and Prospects', *Qualitative Methods*, 2 (2), 3–13.

Berg-Schlosser, Dirk, Gisèle De Meur, Benoît Rihoux and Charles C. Ragin (2009) 'Qualitative Comparative Analysis (QCA) as an Approach', in Benoît Rihoux and Charles. C. Ragin (eds) *Configurational Comparative Methods. Qualitative Comparative Analysis (QCA) and Related Techniques* (Thousand Oaks: Sage), 1–18.

Blatter, Joachim and Till Blume (2008) 'In Search of Co-variance, Causal Mechanisms or Congruence? Towards a Plural Understanding of Case Studies', *Swiss Political Science Review*, 14 (2), 315–56.

Blatter, Joachim and Markus Haverland (2012) *Designing Case Studies* (Houndmills: Palgrave Macmillan).

Brady, Henry and David Collier (eds) (2010) *Rethinking Social Enquiry* (New York: Rowman & Littlefield).

Braumoeller, Bear F. and Gary Goertz (2000) 'The Methodology of Necessary Conditions', *American Journal of Political Science*, 44 (4), 844–58.

Cronqvist, Lasse and Dirk Berg-Schlosser (2009) 'Multi-value QCA (MVQCA)', in Benoît Rihoux and Charles C. Ragin (eds) *Configurational Comparative Methods. Qualitative Comparative Analysis (QCA) and Related Techniques* (Thousand Oaks: Sage).

De Meur, Gisèle and Benoît Rihoux (eds) (2002) *L'analyse quali-quantitative comparée (AQQC-QCA): approche, techniques et applications en sciences humaines* (Louvain-la-Neuve: Academia-Bruylant).

Dion, Douglas (1998) 'Evidence and Inference in the Comparative Case Study', *Comparative Politics*, 30 (2), 127–45.

Emmenegger, Patrick, Jon Kvist and Svend-Erik Skaaning (2013) 'Making the Most of Configurational Comparative Analysis: An Assessment of QCA Applications in Comparative Welfare-State Research', *Political Research Quarterly*, 66 (1), 185–90.

Engeli, Isabelle and Christine Rothmayr (2013) 'Diverging Against all Odds? Regulatory Paths in Embryonic Stem Cell Research Across Western Europe', *Journal of European Public Policy*, 20 (3), 407–24.

Fiss, Peer C., Dmitry Sharapov and Lasse Cronqvist (2013) 'Opposites Attract? Opportunities and Challenges for Integrating Large-N QCA and Econometric Analysis', *Political Research Quarterly*, 66 (1), 191–97.

George, Alexander L. and Andrew Bennett (2005) *Case Studies and Theory Development in the Social Sciences* (Cambridge: MIT Press).

Gerring, John (2007) *Case Study Research: Principles and Practices* (Cambridge University Press).

Grofman, Bernard and Carsten Q. Schneider (2009) 'An Introduction to Crisp Set QCA, with a Comparison to Binary Logistic Regression', *Political Research Quarterly*, 62 (4), 662–72.

Hug, Simon (2012) 'Limitations of QCA', *APSA-Comparative Politics Newsletter*, 22 (1), 6.

Lieberson, Stanley (2004) 'Comments on the Use and Utility of QCA', *Qualitative Methods*, 2, 13–14.

Longest, Kyle and Stephen Vaisey (2008) 'Fuzzy: A Program for Performing Qualitative Comparative Analysis (QCA) in Stata', *The Stata Journal*, 8 (1), 79–104.

Maggetti, Martino and David Levi-Faur (2013) 'Dealing with Errors in QCA', *Political Research Quarterly*, 66 (1), 198–204.

Mahoney, James and Dietrich Rueschemeyer (eds) (2003) *Comparative Historical Analysis in the Social Sciences* (Cambridge University Press).

Marx, Axel (2010) 'Crisp-Set Qualitative Comparative Analysis (csQCA) and Model Specification: Benchmarks for Future csQCA Applications', *International Journal of Multiple Research Approaches*, 4 (2), 138–58.

Marx, Axel and Adrian Dusa (2011) 'Crisp-Set Qualitative Comparative Analysis (csQCA), Contradictions and Consistency Benchmarks for Model Specification', *Methodological Innovations Online*, 6 (2), 103–48.

Marx, Axel, Benoît Rihoux and Charles C. Ragin (2013) 'The Origins, Development and Application of Qualitative Comparative Analysis (QCA): The First 25 Years', *European Political Science Review*, 6 (1), 115–42.

McBride, Dorothy and Amy G. Mazur (2010) *The Politics of State Feminism. Innovation in Comparative Research* (Philadelphia: Temple University Press).

McVeigh, Rory, Carl Neblett and Sarah Shafiq (2006) 'Explaining Social Movement Outcomes: Multiorganizational Fields and Hate Crime Reporting', *Mobilization*, 11, 23–49.

Mello, Patrick A. (2013) 'From Prospect to Practice: A Critical Review of Applications in Fuzzy-Set Qualitative Comparative Analysis', Paper Prepared for Delivery at the 8th Pan-European Conference on International Relations, Warsaw, 18–21 September 2013.

Mill, John Stuart (1967, 1843) *A System of Logic* (University of Toronto Press).Moses, Jonathon W and Torbjorn L. Knutsen (2007) *Ways of Knowing: Competing Methodologies in Social and Political Research* (New York: Palgrave Macmillan).

Moses, Jonathon, Benoît Rihoux and Bernhard Kittel (2005) 'Mapping Political Methodology: Reflections on a European Perspective', *European Political Science*, 4 (1), 55–68.

Ragin, Charles C. (1987) *The Comparative Method. Moving beyond Qualitative and Quantitative Strategies* (Berkeley: University of California Press).

Ragin, Charles C. (1997) 'Turning the Tables: How Case-Oriented Methods Challenge Variable-Oriented Methods', *Comparative Social Research*, 16, 27–42.

Ragin, Charles C. (2000) *Fuzzy-Set Social Science* (Chicago International Press).

Ragin, Charles C. (2006) 'Set Relations in Social Research: Evaluating Their Consistency and Coverage', *Political Analysis*, 14 (3), 291–310.

Ragin, Charles C. (2008a) *Redesigning Social Inquiry: Fuzzy Sets and Beyond* (Chicago University Press).

Ragin, Charles C. (2008b) 'Measurement Versus Calibration: A Set-Theoretic Approach', in Janet Box-Steffensmeier, Henry Brady and David Collier (eds) *The Oxford Handbook of Political Methodology* (Oxford University Press), 174–98. Ragin, Charles C. (2003) 'Comparative Method', in Michael Lewis-Beck, Alan Bryman and Tim Futing Liao (eds) *Encyclopedia of Research Methods for the Social Sciences* (Thousand Oaks: Sage), 148–54.

Ragin, Charles C. and Sean Davey (2012) *Fs/QCA, Version 2.5* (Tucson: University of Arizona).

Ragin, Charles C. and John Sonnett (2004) 'Between Complexity and Parsimony: Limited Diversity, Counterfactual Cases and Comparative Analysis', in Michael Minkenberg and Sabine Kropp (eds) *Vergleichen in der Politikwissenschaft* (Wiesbaden: VS Verlag), 180–97.Rihoux, Benoît and Heike Grimm (2006) 'Beyond the "Qualitative–Quantitative" Divide: Innovative Comparative Methods for Policy Analysis', in Benoît Rihoux and Heike Grimm (eds) *Innovative Comparative Methods for Policy Analysis* (New York: Springer), 1–9.

Rihoux, Benoît and Bojana Lobe (2009) 'The Case for Qualitative Comparative Analysis (QCA): Adding Leverage for Thick Cross-Case Comparison', in David Byrne and Charles C. Ragin (eds) *The Sage Handbook of CaseBased Methods* (London: Sage), 222–43.

Rihoux, Benoît and Charles C. Ragin (eds) (2009) *Configurational Comparative Methods. Qualitative Comparative Analysis (QCA) and Related Techniques* (Thousand Oaks: Sage).

Rihoux, Benoît, Ilona Rezsöhazy and Damien Bol (2011) 'Qualitative Comparative Analysis (QCA) in Public Policy Analysis: An Extensive Review', *German Policy Studies Journal*, 7 (3), 9–82.

Rihoux, Benoît, Priscilla Alamos-Concha, Damien Bol, Axel Marx and Ilona Rezsöhazy (2013) 'From Niche to Mainstream Method? A Comprehensive Mapping of QCA Applications in Journal Articles from 1984 to 2011', *Political Science Quarterly*, 66 (1), 175–84.

Rohlfing, Ingo and Carsten O. Schneider (2013) 'Improving Research on Necessary Conditions: Formalized Case Selection for Process Tracing after QCA', *Political Research Quarterly*, 66 (1), 220–35.

Schneider, Carsten Q. and Claudius Wagemann (2012) *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis* (Cambridge University Press).

Seawright, Jason (2005) 'Qualitative Comparative Analysis Vis-à-vis Regression', *Studies in Comparative International Development*, 40 (1), 3–26.

Sprinz, Detlef F. and Yael Nahmias-Wolinsky (eds) (2004) *Models, Numbers and Cases: Methods for Studying International Relations* (Ann Arbor: University of Michigan Press).

Thiem, Alrik (2010) 'Coverage in Fuzzy Subset Relations', Wolfram Demonstrations Project, http://demonstrations.wolfram.com/CoverageInFuzzySubsetRelations/, date accessed 15 January 2013.

Thiem, Alrik and Adrian Dusa (2013) *Qualitative Comparative Analysis with R: A User's Guide* (New York: Springer).

Varone, Frédéric, Benoît Rihoux and Axel Marc (2006) 'A New Method for Policy Evaluation? Longstanding Challenges and the Possibilities of Qualitative Comparative Analysis (QCA)', in Benoît Rihoux and Heike Grimm (eds) *Innovative Comparative Methods for Policy Analysis* (New York: Springer), 213–36.

Vis, Barbara (2012) 'The Comparative Advantages of fsQCA and Regression Analysis for Moderately Large-N Analyses', *Sociological Methods & Research*, 41 (1), 128–98.

Yin, Robert K. (2008) *Case Study Research: Design and Methods* (Thousand Oaks: Sage).

6

Quantitative Methodologies in Public Policy

Christian Breunig and *John S. Ahlquist*

Introduction

In 2012, American debt rose to over 16 trillion dollars; Greek government spending reached levels nearly 20 percent lower than 2010; official Spanish unemployment crossed the 25 percent mark; French public spending rose to over 55 percent of the GDP; and income inequality (as measured by the Gini Index) in Italy continued to grow. These estimates illustrate the many quantitative measures, descriptions, and analyses in our daily lives.

The comparative study of public policies also relies on assigning numbers to social, economic, and political phenomena with the goal of examining their relationships.[1] Quantitative analysis provides one avenue for understanding policy variations and their political and institutional causes.

The volume of quantitative work in comparative public policy has grown over the last 50 years. Early studies of public policy were predominantly qualitative in nature (see Simeon, 1976 for an early review) because understanding the entire policy process was seen as the central goal in the public policy literature. From early on, researchers were astutely aware of the complexity of this process. Starting in the 1970s, scholars have broadened their theoretical and empirical inquiries across political entities and time and even across policy domains. This expansion encouraged political scientists to develop and employ methodological tools to summarize, describe, and draw conclusions from large amounts of data. Beginning with the 1990s, we have witnessed a marked sophistication in our methodological tools. Hand in hand with these improvements emerged vigorous discussions about the appropriateness and utility of the newfound tools.

   With this brief historical evolution in mind, we introduce three basic goals of quantitative methodology: description, theory testing, and prediction. These three goals are introduced in more detail in the following pages. We illustrate each goal using data and particular quantitative procedures that are frequently employed in the comparative study of the welfare state in advanced democracies. This example is especially valuable here because the welfare state is one of the most researched areas of comparative public policy and one of the most 'quantified' areas of public policy. We specify the relevant data sets when we explore how quantitative methods allow us to describe the welfare state.

   Almost all information about the welfare state and its analysis are presented visually. Public policy scholars quantify many political phenomena and collect incredibly detailed data. However, it remains challenging to present this richness in comprehensible and interesting ways. Visualization helps researchers and their consumers to explore, compare, and understand data (see Tukey, 1977; Tufte, 1983). We believe that well-crafted graphical displays allow quantitative scholars to tell their story more convincingly and clearly. This is particularly true for describing outputs from statistical analysis and simulations. Visual displays enable audiences who have no background in the complexities of the model or estimation technique to grasp intuitively the main findings of a research project.

   Given the limited scope of this chapter, several themes in methodological debates remain untouched. First, we do not dissect and evaluate differences between qualitative and quantitative methodologies in comparative public policy. Instead, we simply believe that description and inferential leverage are valid and realizable ambitions for both types of research. Second, and related, we do not focus on the many important and difficult decisions that go into an effective research design; we simply assume that the basic components of a successful design, such as comparison and control, are given. Third, this chapter does not touch upon other areas of public policy research that often rely on quantitative methodology. Among them are important fields of public administration research such as evaluation research and cost-benefit analysis. A typical tool of the first area is survey research. Since the 2000s, evaluation research more regularly incorporates clear experimental design. We discuss the role of experiments in the section on inference but do not explicitly refer to evaluation research. The second area, cost-benefit analysis, is often associated with the need to monetize various policy alternatives as well as their potential costs and impacts in order to develop efficient policy choices. It is also outside the realm of this

chapter because comparative public policy research in political science is seldom directed at explaining the effectiveness and efficiency of a particular policy. To put it simply, we focus on quantitative analyses of comparative policy outcomes.

Measurement and data

Public policy scholars routinely develop concepts that are not directly observable. It is, therefore, a fundamental task to develop procedures that enable us to assign numbers to events or other empirical objects such as institutional structures, control of government, partisan ideology, or the welfare state. This process is called measurement (see, for example, Frankfort-Nachmias and Nachmias, 2008; Jackman, 2008). In this section, we introduce two components of measurement: levels of measurement and assessments of measurement error (that is, validity and reliability). It should be clear that both level and quality of measurement have important consequences for our ability to describe and assess relationships among political phenomena. This section also introduces different types of data structure as well as the data sets employed in the subsequent analysis of welfare states.

A perennial concern of measurement relates to the quality, defined in terms of accuracy and consistency, of the measurement. Measurement error doesn't stem from real differences in the properties of an object but is a consequence of variations in the measurement procedure itself. The two main concerns of measurement are reliability and validity. Reliability describes the extent to which a measurement is consistent under the same procedures. Validity is the idea that researchers are measuring what they intend to measure. In order to validate a measurement, researchers need to supply supporting evidence for its appropriateness. Three types of validity are important here: content validity assesses to what degree the instrument adequately captures all characteristics of a concept; empirical validity captures the degree to which concept and measurement are consistently related. Construct validity indicates to what extent the measurement connects to a broader theoretical framework.

Questions of measurement feature prominently in welfare state studies. In particular, Esping-Andersen's (1990) path-breaking work as well as studies of retrenchment (Pierson, 1996; Korpi and Palme, 2003) highlight how debates about concept formation, operationalization, and measurement can evolve into central concerns of the field. Before Esping-Andersen objected that citizens do not struggle for spending per se (Esping-Andersen, 1990: p. 21), the most common measure of the

welfare state was public social expenditures as a percent of the gross domestic product (GDP). Instead, Esping-Andersen stresses the importance of entitlements for the elderly, sick, and unemployed. His argument recasts the welfare state in terms of social insurance and social services. Similarly, Pierson argues that expenditure measures barely pick up the subtle changes in the welfare state that lead to retrenchment.

The ensuing discussion on the 'dependent variable problem' (see also Green-Pedersen, 2004) highlights the importance of the three types of measurement in quantitative public policy work. First, the retrenchment literature challenges the content validity of the public expenditure measure. In contrast to the narrow focus on total public or social expenditures, programmatic measures that capture the insurance motive of the welfare state were missing or incomplete. Instead of simple government activity on spending, more recent works rely on more concrete individual level benefits (such as unemployment benefit replacement rates) that better capture the decommodifying role of the welfare state. Second, retrenchment scholars contest the empirical validity of the spending measure. They claim that expenditure is a poor instrument for measuring outcomes by showing that restrictions on entitlements are not picked up by the spending variable. They also argue that much of the inclusive evidence with regard to major theoretical expectations, such as partisan ideology or trade openness, at least partly, are due to using government spending as a dependent variable.

For the purpose of this chapter, we simply acknowledge that there is little consensus on how exactly one can capture and operationalize the nature of the welfare state and its scope. We therefore rely on two common measures. First, we use public social expenditures (as percent of GDP) supplied by Busemeyer (2009) as the 'classic' measure of government welfare state effort. Second, we employ the welfare generosity index constructed by Allen and Scruggs as described in Scruggs (2006). The index follows Esping-Andersen and measures the generosity of welfare benefits at the individual level. It combines replacement and coverage rates for three core social insurance programmes: unemployment, sickness, and pensions. The index ranges from 0 (no benefits) to 100 (total replacement of income and coverage). Both measures cover the two competing conceptualizations of the welfare state.

Our data is structured as time-series cross-section (TSCS). This type of data holds repeated observations of a set of units (here countries) over a period of time and is common in comparative public policy. One advantage of pooling data as a TSCS sample is that researchers can learn from both the time as well as the cross-sectional component of the data.

This type of analysis is at the heart of comparative politics and public policy. However, special care needs to be taken with data of this sort because extensive correlation across observations, both longitudinally and in space, can prevail. This is especially true when relying on longer time series for a dozen or two countries. The sample of this study consists of 18 Organization for Economic Co-operation and Development (OECD) member states over 22 years (1980–2002), which results in 396 cases.

TSCS data are essentially a combination of two other types of data structures that public policy scholars regularly encounter. The most basic data is a collection of observations at one point in time. This type of data is called cross-sectional. Typical examples of cross-sectional data are surveys or experiments. Time-series data are measurements of the same unit over time. In addition, many policy scholars are interested in policy diffusion processes. Diffusion scholars are interested in the spatial nature of the data; that is, the units of observation are distributed in space either as regions or point processes (see Chapter 9 in this book). Clearly, some data are combinations of these structures and require special care in accounting for the complex dependencies among observations.

Objectives of quantitative analysis in comparative public policy

Quantitative analysis in political science and public policy concerns three core objectives. First, description enables researchers to illustrate and summarize large amounts of data intuitively and efficiently. Second, causal inference and theory testing allows researchers to state and estimate relationships among variables. It also provides measures of uncertainty for those relationships. The third objective is predicting outcomes based on data and models. At the centre of all three objectives lies the regression model. It is the workhorse for quantitative analysis and is utilized for all three aims.

Description

Both quantitative and qualitative work in comparative public policy engage in describing political phenomena using numbers. Descriptions of single variables usually summarize key properties of the distribution of the interested observations. Typical statistics are measures of central tendencies and dispersion such as means and standard deviations. Similarly, visual techniques such as histograms or box plots can effectively illustrate large amounts of data. For comparative public policy work, line plots that display the development of particular policy outcomes

across time or space are especially important. Description and descriptive statistics summarize data in effective and intuitive ways. Essentially, they are a tool for reducing information in order to gain understanding of the subject at hand.

Figure 6.1 describes the welfare state data. The top row combines all data into a histogram and the bottom row displays country specific developments over time. On average, the countries in our sample spend 21 percent of GDP on public social expenditures. The welfare effort ranges from 9 percent of GDP in 1980 Greece to 36 percent in 1993 Sweden; 50 percent of all observations lie between 17 and 26 percent. The generosity index, which is the product of replacement rate and coverage rate summed over three basic social insurance programmes, is
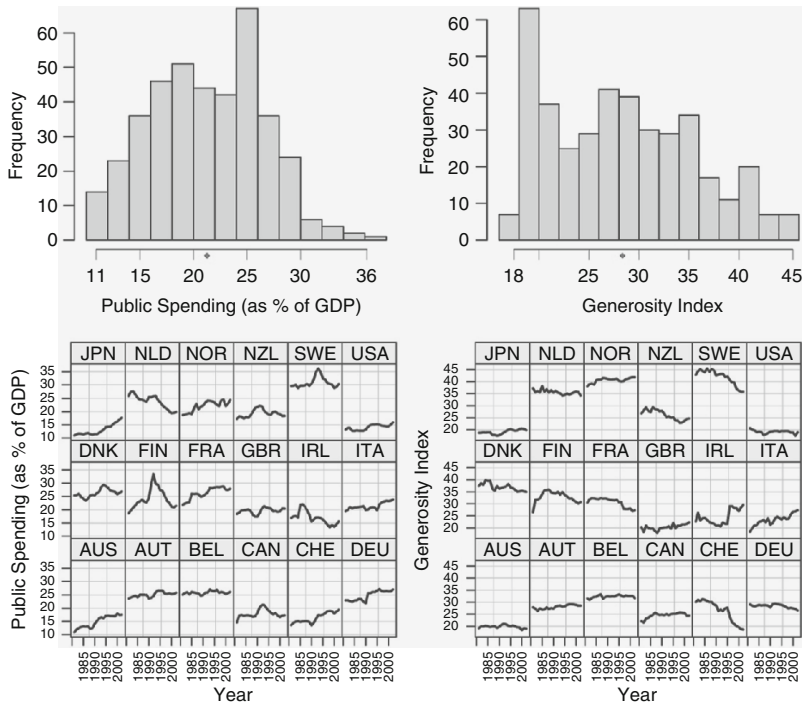
Figure 6.1   Description of the welfare state
The left side of the plot displays measures of public spending and the right side concentrates on the generosity index. The top row presents histograms. The *x*-axis provides information about the range, interquartile range, mean (grey dot), and median of the variable. The bottom row uses line plots to show how the welfare state changes over time in each country.

a bit harder to interpret. The average score is 28, which corresponds to Germany and France's scores in the late 1990s and ranges between 18 (Japan 1980) and 45 (Sweden in the late 1980s).

The bottom row of Figure 6.1 displays the development of the welfare state across countries and time. The line plots strikingly show clear cross-country differences in terms of public spending and generosity. They also reveal substantive variation across time in several countries such as Finland, Italy, and Switzerland.

The side by side comparison of public spending and the generosity index also hint that the two measures of the welfare state do not correlate perfectly. A simple tool for examining the association between two variables is the scatter plot. Figure 6.2 displays both measures. It appears that there is positive relationship between public spending and generosity. Generosity increases jointly with welfare state spending. The plot also suggests that, at high levels of generosity, the spread of public spending is quite large. One reason for this dispersion might be that, despite generous benefits, the state does not necessarily have to make huge spending commitments. This is probably true for periods of economic expansion.

Interesting outcomes in comparative public policy, just like other social phenomena, almost always have more than one cause. In order to
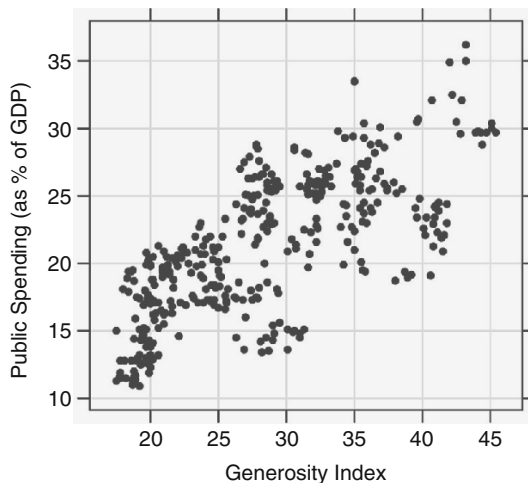
Figure 6.2 Relationship between two measurements of the welfare state
The scatter plot displays the correlation between public spending and the generosity index for 18 OECD countries, 1980–2002.

isolate and assess the theoretical expectations that are most relevant to a research agenda, we need to account and control for alternative factors. We use multiple regression in observational studies in order to accomplish this task.

Inference and theory testing

The regression model and its extensions can be employed as a descriptive tool for analysing empirical regularities among variables. The more important concern in comparative public policy is to identify and assess causal relationships. Among the different approaches to causality in the social sciences (see Brady, 2008), the idea of counterfactuals are often employed in comparative public policy. For example, what would have been the environmental standards in Italy had it not been part of the European Union? Would education at the high-school level be better if the government implemented nation-wide curriculum requirements? However, it is impossible to simultaneously observe Italy's environmental standards both with and without the role of the European Union. This obstacle is known as the fundamental problem of causal inference. Since we cannot rerun history, we need to rely on repeated measures and then separate the variable of interest ('treatment') for all other confounding variables. Doing this allows us to estimate a causal effect.

The most effective method for identifying causal effects is a randomized assignment of a treatment to the units. However, random assignments and experimental designs as well as natural experiments are rare or impossible in comparative public policy. Therefore quantitative scholars using observational data need to be cautious about asserting causal relationships. In recent years, matching models have been employed in order to assess causal claims. Matching tools group observations that are as similar as possible along all dimensions except for the variable of interest and then compute differences in the outcome in order to establish causal effects.

The most common approach is to rely on a multiple regression model for theory testing. Typically, a researcher develops some theoretical expectations about relationships among variables and formulates them as distinct and testable hypotheses. Quantitative methods then allow describing whether the theoretical expectations are consistent with the sample of their data. Based on their analysis, researchers are in a position to quantify their uncertainty about the estimated relationships. It is important to point out that these statements of uncertainty – often expressed in terms of statistical significance or standard errors – are derived for statistical assumptions.

The study of the welfare state has been a recurring subject in the scholarly debate about the appropriate use of multiple regression and its challenges. We follow Busemeyer's review on the link between globalization and the welfare state. Two competing hypotheses about the relationship between globalization (that is, economic openness) and the welfare state have been proposed. The competition thesis holds that two mechanisms force governments to decrease public spending as economic globalization increases. First, the exit threat of mobile capital forces governments to lower their taxes and thereby reduce the room for expenditures. Second, international financial markets punish countries that rely on deficit spending. In short, the relationship between economic openness and welfare state spending should be negative. In contrast, the compensation thesis posits that government spending increases as economic openness progresses. Governments are asked to cushion the effects of globalization on domestic labour and employment. Some authors go so far as to argue that increased spending, for example on retraining and education, can in fact contribute to the competitiveness in international markets.

With this theoretical starting point in mind, we would like to elaborate on two 'relatively easy' problems of causal inference in a multiple regression framework. The discussion also highlights some common statistical assumptions. The first concern centres on model specification. Since the social world is complex and potentially many variables might contribute to explaining an interesting empirical regularity in public policy, researchers need to settle on a regression model that they would like to analyse. By selecting what variables to include or exclude in the equation they might run into two issues: inclusion of irrelevant and omission of relevant variables. Both have consequences for the quality of the estimation.

The problem of 'too many' variables is relatively subtle. First, more complicated models sacrifice parsimony and, therefore, become harder to understand and use. Second, more complicated models make it unlikely that all possible specifications can be effectively presented (see Achen, 2002). Third, on a technical level, adding variables can increase our uncertainty about the estimates for the other covariates due to problems of highly correlated regressors. Adding too many variables might lead researchers to miss some systematic relationships.

The generally more severe problem occurs when researchers omit relevant variables. In this situation, our model estimates can be biased, perhaps badly so. In other words, a researcher might over- or under-estimate the included variables' effects (depending on the correlation

between the excluded and included variables). Even worse, the downward bias of the standard errors implies that one might overstate the influence of the included variables. Given this problem, researchers in comparative public policy need to think carefully about rival hypothesis and model them. The inability to find a measure for an important concept does not lessen this problem.

For replication and illustration purposes, we follow Busemeyer's model selection. He argues that the compensation thesis contains a static and dynamic aspect. The static argument purports that levels of globalization are associated with spending and that there is a qualitative difference between globalization at its onset in 1980 and in its later state in 2000. This argument is essentially based on Katzenstein's (1985) comparison. Consequently, we estimate the following model, including several factors that might shore up the demand for welfare.

$$
\begin{aligned}
Welfare_{it} = {} & \alpha + \beta_1 Openness_{it} + \beta_2 Unemployment_{it} + \beta_3 Income_{it} \\
& + \beta_4 Age\ Dependency_{it} + \beta_5 Deindustrialization_{it} \\
& + \beta_6 Female\ Labor_{it} + \varepsilon
\end{aligned}
\tag{1}
$$

While equation (1) is estimated as a cross-sectional regression in 1980, 1990, and 2000, the dynamic feature of globalization suggests that the relationship is really between changes in openness and changes in spending. This specification attempts to capture the increasing process of economic liberalization. In order to test this logic, we analyse the following first-difference model on the TSCS data.

$$
\begin{aligned}
\Delta Welfare_{it} = {} & \alpha + \beta_1 Welfare_{it-1} + \beta_2 \Delta Openness_{it} + \beta_3 \Delta Unemployment_{it} \\
& + \beta_4 \Delta Income_{it} + \beta_5 \Delta Age\ Dependency_{it} + \beta_6 \Delta Inflation_{it} \\
& + \beta_7 \Delta Deindustrialization_{it} + \beta_8 Maastricht_{it} \\
& + \beta_9 Time_{it} + \varepsilon
\end{aligned}
\tag{2}
$$

For our example, we evaluate the compensation thesis by controlling for a number of alternative explanations that might affect the demand for more welfare. In particular, we consider the following controls. Unemployment is captured by the commonly used definition of the OECD unemployment rate. Income per capita is measured as current US dollar PPP. Age dependency ratio divides the population share of 0–14 and over 65 years old by the population share of the 15–65 years old. Deindustrialization is computed by subtracting the share of agricultural and industry employment from 100. The female labour force is measured as working women in percentage of population from 15–64 years. Economic openness is measured as the average between exports and imports as a percentage of GDP. Inflation is measured as the annual

changes in consumer price indices. The Maastricht variable indicates EU membership after 1992. All data come from Busemeyer (2009).

The second common problem of multiple regression in comparative public policy is the need to properly account for complicated data structures accompanied by relative small sample sizes. In particular, TSCS data demands that special care be taken when modelling the longitudinal and the spatial dependencies appropriately. In a typical application, researchers observe annual data in a number of advanced democracies. For comparative public policy, it is also often the case that the number of units is smaller than the number of years in the sample. In an influential essay (the most cited article in *American Political Science Review*), Beck and Katz (1995) brought modelling issues of TSCS to a large audience. The relative ease in implementing their recommendations contributed to an explosion of TSCS analyses since the 1990s. In a critical review of political science literature on the issue, Wilson and Butler (2007) conclude that:

> It is more than a little ironic that even though B and K's analysis focused on the danger of using estimators without fully understanding their properties, so many in the profession applied the B and K method without paying any attention to the simple textbook issues.

Their review suggests that sufficient methodological knowledge is required in order to grasp the complications of TSCS data (Wooldridge, 2010, is a good start). As a consequence, comparative public policy scholars feel the need to not only follow substantive innovations in the field but also pay attention to methodological developments. For example, TSCS data structures might also be understood as multilevel or spatial data (the 2007 special issue in *Political Analysis* 15 (2) is instructive here) and a Bayesian framework might also be utilized (see Western and Jackman, 1994 for an early example). For our illustrations, we follow Busemeyer and employ panel corrected standard errors in order to abate panel heteroskedasticity and contemporaneous correlation (for example, induced by common international shocks).

As we mentioned before, interpreting regression results from tables is a challenging task for lay readers. Public policy scholars can use simulation and visual aids in order to present their results intuitively and effectively. Following King et al. (2000) (also Williams and Whitten, 2012), we use our estimates to produce counterfactual simulations and then display their conditional first difference and 90 percent confidence bands (Adolph, 2012). In the following graphs, we consider the counterfactual

of a 0.5 standard deviation increase in an independent variable. For example, we ask what would happen to welfare state spending if we increase trade openness from 33 percent of the GDP to 42 percent of the GDP. In each plot, the dots are the point estimates for the counterfactual and the lines are the 90 percent confidence bands. If a confidence band does not cross zero, we are confident that our estimate is different from zero (in other words statistically significant).

Figure 6.3 presents the results from the cross-sectional analyses. For the years 1980, 1990, and 2000, we estimated the determinants of the welfare state. The left plot displays the results for the spending measure and the right plot shows the results for the generosity index. The right graph shows that trade openness increases public social spending in all three cross-sections. For 1980 and 1990, our estimates are clearly different from zero (that is, the confidence band does not touch the vertical zero line). Moreover, the size of the estimated effect is substantial. A 0.5 standard deviation increase (about 9 percent of GDP) leads to an increase of about 1.5 percent of GDP in public social spending, *ceteris paribus*. By 2000, the effect of trade openness on spending is no longer distinguishable from zero. It is also noteworthy that all other covariates in all three cross-sections are not statistically significant. The spending results are mirrored when we measure the welfare state in terms of individual levels of generosity. As the left plot of Figure 6.3 illustrates, a 0.5 standard deviation increase in trade openness leads to a ca. 2 point increase in generosity in 2000 and at least a 3 point increase in 1980 and 1990, holding everything else equal. Roughly speaking, these results translate into the difference between the Danish and the French generosity levels during the mid-1980s. Taken together, we can posit that we cannot reject the compensation hypothesis in the cross-sectional data. Increases in trade openness are associated with higher welfare state efforts in 1980, 1990, and 2000.

Figure 6.4 assesses the dynamic logic of the compensation hypothesis using data from 18 OECD countries for the period between 1980 and 2002. The top plot illustrates the estimated effects on changes in public social spending and the bottom shows the results for changes in the generosity index. Concentrating on spending first, we find that trade openness has a substantive and negative effect on changes in public social spending. A 0.5 standard deviation increase in the change of trade openness (ca. 1 percent of GDP) leads to a 0.7 percent of GDP reduction in public social spending *ceteris paribus*. Since the confidence band does not touch the zero line, the estimated effect is statistically different from zero. This implies that positive changes in trade openness lead to welfare state cutbacks. In addition to trade, changes in unemployment
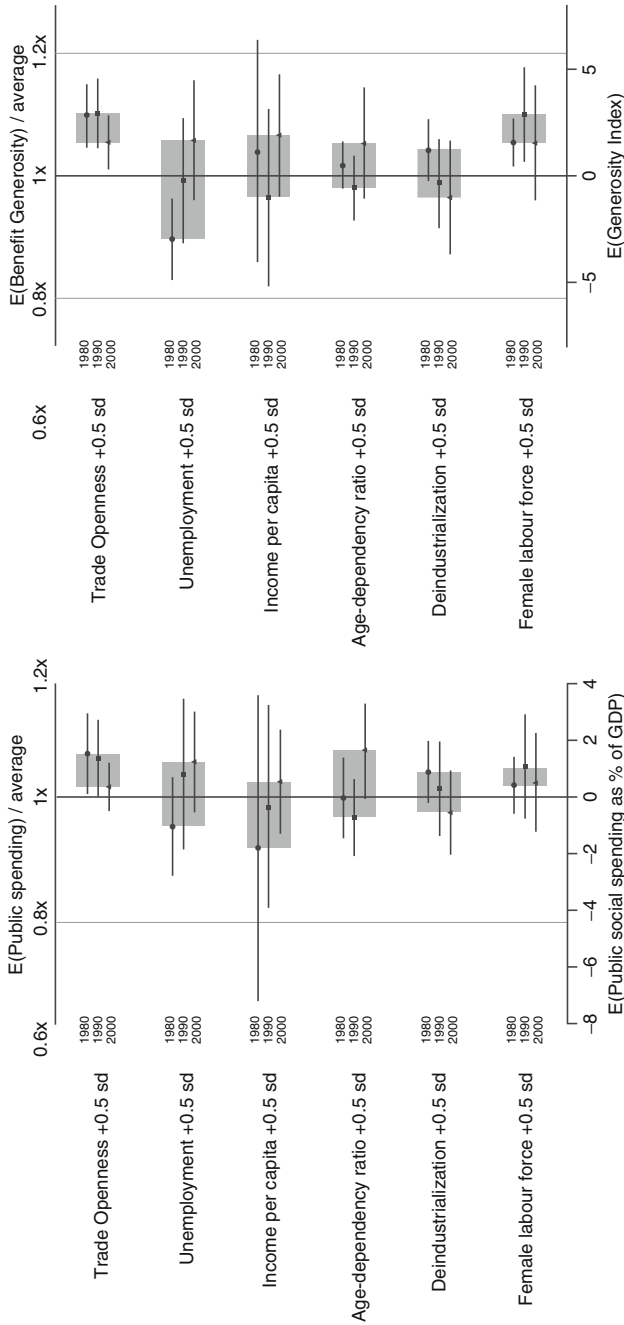
Figure 6.3 Cross-sectional analysis of the welfare state
The plots display the simulated effect for public social spending (left) and the generosity index (right) for each counterfactual listed on the left side of each plot. The dots are the point estimates. Shaded grey boxes show the full range of the point estimates. The lines are the 90 confidence bands. The bottom *x*-axis displays the expected changes in the welfare state effort.

Change in Trade Openness +0.5 sd
Change in Unemployment +0.5 sd
Change in Income per capita +0.5 sd
Change in Age-dependency ratio +0.5 sd
Change in Deindustrialization +0.5 sd
Change in Inflation +0.5 sd

−0.2 −0.1   0   0.1   0.2

E(Change in Public social spending)
(as % of GDP)

Change in Trade Openness +0.5 sd
Change in Unemployment +0.5 sd
Change in Income per capita +0.5 sd
Change in Age-dependency ratio +0.5 sd
Change in Deindustrialization +0.5 sd
Change in Inflation +0.5 sd

−0.15 −0.1 −0.05 0 0.05 0.1 0.15

E(Change in Generosity Index)

Figure 6.4   Time-series cross-sectional analysis of the welfare state
The plots display the estimated effects for change in public social spending (top)
and change in the generosity index (bottom). In each plot, the counterfactual is
listed on the left side. The plot displays the point estimate and the confidence band.

and income also display statistically significant effects. While more
unemployment leads to spending increases, the growth of personal
income reduces welfare state spending. In contrast to the spending
measurement, the results for changes in the generosity index (bottom
of Figure 6.4) do not display any statistically significant effects. The
direction of the trade openness variable is negative. Increases in the
change of trade openness lead to reductions in generosity.

Some possible conclusions from the multiple regressions are: first and
substantively, the ability of the welfare state to compensate for global-
ization initially existed, but its effect diminished between 1980 and
2000 as the cross-sectional regressions show. The TSCS analysis indicates
that increased openness leads to reductions in the welfare state (measured
in spending) across countries and time. Second, the regressions reveal

that careful modelling of theoretical logic can provide researchers with important and varied insights using the same data set. Third, the difference in the results between our two measurements of welfare states in the TSCS analysis indicate that careful measurement and conceptualization remains a crucial task for quantitative public policy research. Fourth, visual displays can make the outputs of statistical and simulation results more easily understandable and comparable. For example, differences in the cross-sectional results are easily captured in Figure 6.3. Given this, visualization and the advancement of their techniques is an important task for comparative public policy scholars.

Prediction

Prediction is the third goal of quantitative analysis. In contrast to predictions about electoral outcomes and wars, this tool hasn't been sufficiently exploited in comparative public policy. In the most common form, researchers rely on regression models in order to develop their predictions. When we speak of prediction we usually mean forecasting, that is, calculating unrealized events into the future. For comparative public policy scholars, forecasting can serve as a helpful tool for illustrating policy consequences of their theoretical models and for making recommendations about policy alternatives.

One other predictive tool might be equally important. Out-of-sample prediction can be employed for assessing how well a model estimated on one set of data is able to predict another set of data not used in the estimation. In addition, out-of-sample prediction (or cross-validation) enables researchers to assess their results in substantively important terms.

The beauty of this method can be illustrated using our welfare data set. We estimate the second model for changes in public social spending. Now we can ask ourselves, 'Which countries do not fit the model well?' In order to answer this question, we leave one country out of the estimation and then compare the model's predictions for that country with the actual observations. The difference between the 'forecast' and the actual observations can be quantified in terms of mean absolute forecast error. This measure has the same units as our original dependent variable – change in public social spending as a percent of GDP.

Figure 6.5 displays the results of this exercise for each country. The larger the mean absolute error, the less well our specification fits a country compared to all the others. For our dynamic TSCS model of changes in public social spending, the model seems to work well for liberal welfare states such as the United States and Japan. On the other hand, Finland and Ireland have the largest 'forecasting errors'. For example,
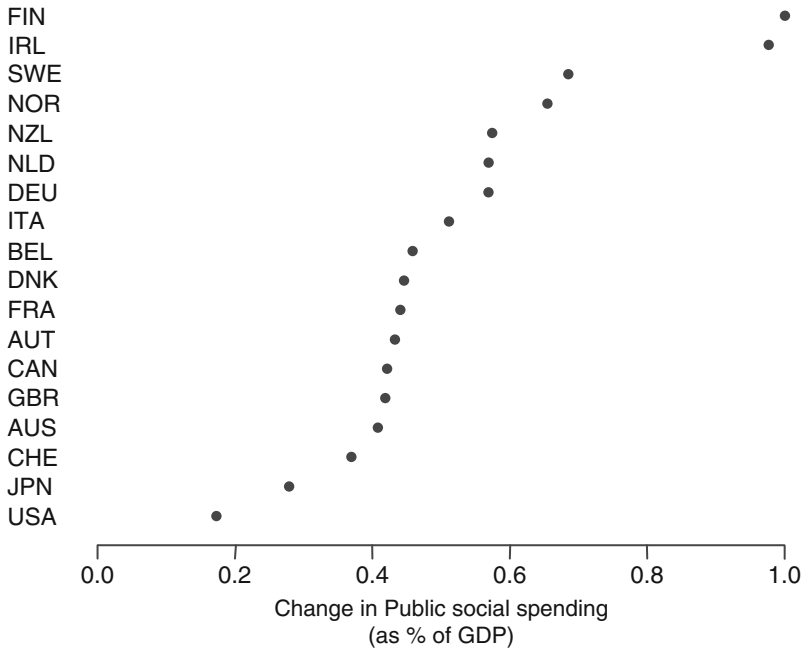
Figure 6.5   Mean absolute error by countries
The results are based on a TSCS regression in the 18 OECD countries, 1980–2002.

the mean forecast error for Finland is over 1 percent point change in spending (as percent of GDP). This suggests that the Finnish and Irish experience of welfare retrenchment since the 1980s is quite different from that of the other countries. A qualitative examination of these two experiences might explore this divergence.

As a note of caution, the link between 'accurate' predictions and a correct causal model is not straightforward. While appropriate causal models will result in precise predictions, one can also construct models only with the purpose of improving predictive ability in mind. Likewise, it is possible to remove statistically significant variables from a model without degrading the predictive accuracy of the overall model very much.

The Interplay of description and inference: Typology formation and classification

We examine the interplay between descriptive and inferential goals of quantitative analysis in the context of typology formation and classification

(Elman, 2005; Ahlquist and Breunig, 2012). One of the essential tasks of social science research is to define, measure, and then classify objects along several dimensions. The classification of objects captures distinct 'types' of a phenomenon of interest. However, it is equally plausible that researchers have developed theoretical expectations and then examine data in order to identify typologies 'in the real world'. This task is clearly inferential.

Traditional cluster analysis is the typical tool for identifying groups in data. The goal of cluster analysis is to minimize differences among objects within groups and maximize difference among groups. Traditional cluster models can serve as an exploratory and descriptive tool. In particular, there is no foundation in statistical theory on which clustering solution in terms of cluster numbers and groupings should be preferred, no measure of uncertainty about the placement of an object, and only a limited number of geometrical arrangements allowed. In short, substantive interpretation of the results is the responsibility of the analyst.

Model-based clustering using mixture models provides researchers with an inferential tool (Banfield and Raftery, 1993; Fraley and Raftery, 1998). It permits evaluation of uncertainty, introduces a statistical yardstick for model selection, and provides guidance on which variable to use for the analysis. So far, mixture models have received little attention in comparative public policy. We use our welfare state data in order to demonstrate the potential payoffs.

Esping-Andersen's (1990) development of three welfare regimes led to a subsequent effort in identifying welfare state clustering and adding new types – such as the Mediterranean, Latin American, or East Asian welfare state. Using different variables, time periods, and methods, welfare state scholars are still debating on how many worlds of welfare there are, placements of various countries, and changes in regime types over time. The divergent results and findings are at least partly due to the individual researcher's use of exploratory and descriptive methods. It is important to stress that country clusters are based on clear theoretical arguments. As such, the quantitative analysis should be understood as an inferential task.

Will we find welfare state clusters if we rely on the presented measurements of welfare state and model-based clustering? Figure 6.6 displays the results from model-based clustering using public social spending and the generosity index for 18 OECD countries in 1980, 1990, and 2000. The model with the best fit for the data, (that is, the highest Bayesian Information Criterion or BIC), identifies two components with varying variance and ellipsoidal shape. The middle plot displays the estimated density contours. The typical liberal welfare states such as the United States
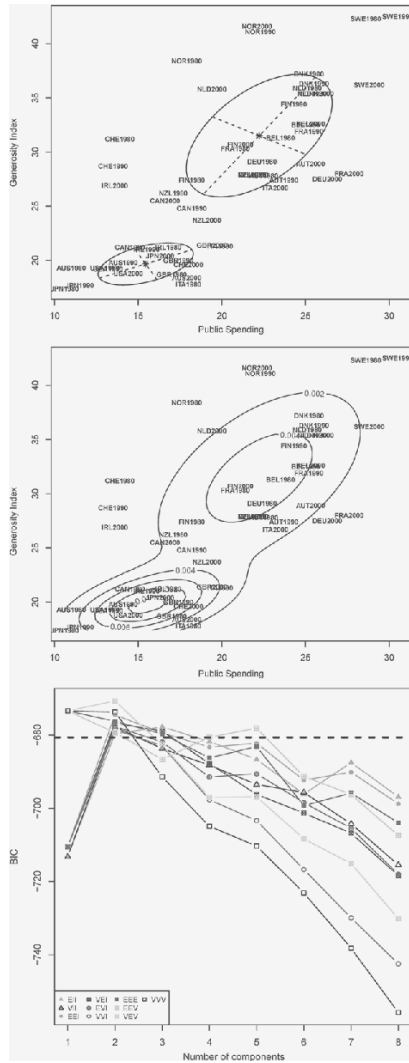
Figure 6.6   Model-based clustering of welfare states
On the top two plots we used country and year labels. The top plot displays the
estimated mixture components. The ellipses in the top subfigure are based on
the estimated mean and variance parameters for the mixture components. The
middle subplot shows the contours of the estimated density surface. The bottom
plot guides model selection. It should be the BIC for various models and number
of components. The broken line represents a BIC value ten less than the BIC of
the best-fitting model.

and Australia are clearly grouped together. The second cluster groups both continental and Nordic countries together. The model suggests that Canada and Ireland moved out of the liberal and into the second cluster by 2000. In addition, there is some uncertainty about the placement of Italy. The probability for being in the second cluster with the other European countries is 0.89 in 1980 and 0.77 in 1990. Similarly, this model indicates that the placement of New Zealand and Great Britain is difficult (low probabilities) in 2000. Finally, the bottom plot indicates the BIC of various models, their shape, and the number of components. While a two cluster solution is the best fitting model, several alternative models ranging from one to five components are not substantively worse. We would be more confident in the two cluster solution if the BIC of all other models were below the broken line (that is, at least ten units lower than the best fitting model). Given this uncertainty, we would caution analysts from proclaiming clear country grouping based on the presented data.

Discussion

We highlighted the three objectives of quantitative analysis in comparative public policy. In order to move from the goal of description to inference, analysts are required to be well versed in the substantive, theoretical, and methodological literature. The welfare state literature is a prime example for this connection. As the substantive understanding of welfare states and its theoretical development matured, important discussions about methodological issues such as measurement, causation, and proper modelling of the intricate data structure took centre stage. In short, the dialogue between theory and methodology contributes to a better understanding of the welfare state and comparative public policy more generally.

Second, the development of quantitative methodologies and their usage in comparative public policy enables us to expand our knowledge and to probe new terrains of research. Given the increasing availability and ease of access to quantitative information, analysts need to be mindful of the origin of the data and the incentives of their creators, such as international organizations, national governments, or public and private interests (see, for example, Soroka et al., 2006; Jerven, 2013). At the same time, we need to be cognizant of vast new data sources such as spatial information from satellite images or Internet traffic patterns. Using, analysing, interpreting, and presenting data remains a worthy challenge.

Finally, we like to close by drawing attention to the closing gap between quantitative and qualitative work. Both methodologies serve similar goals. Qualitative works often display and rely on numerical data and quantitative analysis requires appropriate substantive interpretation. For example, one of the frontiers of quantitative work is the analysis of the written or spoken word. The use of qualitative knowledge can also be directly incorporated into quantitative analysis especially in a Bayesian setting.

Note

1   This chapter builds on on Ahlquist, 2010.

Bibliography

Achen, Christopher H. (2002) 'Towards a New Political Methodology: Microfoundations and ART', *Annual Review of Political Science*, 5 (1), 423–50.

Adolph, Christopher (2012) Tile - R Package, http://faculty.washington.edu/cadolph/software/tile_0.4.7.zip, date accessed 1 June 2012.

Ahlquist, John S. (2010) 'Quantitative Methodologies', in *The Encyclopedia of Political Science*. New York: CQ Press.

Ahlquist, John S. and Christian Breunig (2012) 'Model-Based Clustering and Typologies in the Social Sciences', *Political Analysis*, 20 (1), 92–112.

Banfield, Jeffrey D. and Adrian E. Raftery (1993) 'Model-Based Gaussian and Non-Gaussian Clustering', *Biometrics*, 49 (3), 803–21.

Beck, Nathaniel and Jonathan N. Katz (1995) 'What to Do (and Not to Do) with Time-Series Cross-Section Data', *The American Political Science Review*, 89 (3), 634–47.

Brady, Henry E. (2008) 'Causation and Explanation in Social Science', in Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier (eds) *The Oxford Handbook of Political Methodology* (Oxford University Press), 217–71.

Busemeyer, Marius R. (2009) 'From Myth to Reality: Globalisation and Public Spending in OECD Countries Revisited', *European Journal of Political Research*, 48 (4), 455–82.

Elman, Colin (2005) 'Explanatory Typologies in Qualitative Studies of International Politics', *International Organization*, 59 (2), 293–326.

Esping-Andersen, Gøsta (1990) *The Three Worlds of Welfare Capitalism* (Cambridge: Polity Press).

Fraley, Chris and Adrian E. Raftery (1998) 'How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis', *The Computer Journal*, 41 (8), 578–88.

Frankfort-Nachmias, Chava and David Nachmias (2008) *Research Methods in the Social Sciences* (New York: Macmillan Higher Education).

Green-Pedersen, Christoffer (2004) 'The Dependent Variable Problem Within the Study of Welfare State Retrenchment: Defining the Problem and Looking for Solutions', *Journal of Comparative Policy Analysis: Research and Practice*, 6 (1), 3–14.

Jackman, Simon (2008) 'Measurement', in Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier (eds) *The Oxford Handbook of Political Methodology* (Oxford University Press), 119–52.

Jerven, Morten (2013) *Poor Numbers: How We Are Misled by African Development Statistics and What to Do About It* (Ithaca: Cornell University Press).

Katzenstein, Peter J. (1985) *Small States in World Markets: Industrial Policy in Europe* (Ithaca: Cornell University Press).

King, Gary, Michael Tomz, and Jason Wittenberg (2000) 'Making the Most of Statistical Analyses: Improving Interpretation and Presentation', *American Journal of Political Science*, 44 (2), 347–61.

Korpi, Walter and Joakim Palme (2003) 'New Politics and Class Politics in the Context of Austerity and Globalization: Welfare State Regress in 18 Countries, 1975–95', *American Political Science Review*, 97 (3), 425–46.

Pierson, Paul (1996) 'The New Politics of the Welfare State', *World Politics*, 48 (2), 143–79.

Scruggs, Lyle (2006) 'The Generosity of Social Insurance, 1971–2002', *Oxford Review of Economic Policy*, 22 (3), 349–64.

Simeon, Richard (1976) 'Studying Public Policy', *Canadian Journal of Political Science*, 9 (4), 548–80.

Soroka, Stuart N., Christopher Wlezien, and Iain McLean (2006) 'Public Expenditure in the UK: How Measures Matter', *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169 (2), 255–71.

Tufte, Edward (1983) *The Visual Display of Quantitative Information* (Cheshire: Graphics Press).

Tukey, John (1977) *Exploratory Data Analysis* (Reading: Addison-Wesley).

Western, Bruce and Simon Jackman (1994) 'Bayesian Inference for Comparative Research', *The American Political Science Review*, 88 (2), 412–23.

Williams, Laron K. and Guy D. Whitten (2012) 'But Wait, There's More! Maximizing Substantive Inferences from TSCS Models', *The Journal of Politics*, 74 (3), 685–93.

Wilson, Sven E. and Daniel M. Butler (2007) 'A Lot More to Do: The Sensitivity of Time-Series Cross-Section Analyses to Simple Alternative Specifications', *Political Analysis*, 15 (2), 101–23.

Wooldridge, Jeffrey M. (2010) *Econometric Analysis of Cross Section and Panel Data*, 2nd edn (Boston: The MIT Press).

7
Interpretive Analysis and Comparative Research

Dvora Yanow

Imagine a comparative analysis of primary school breakfast programmes, an undertaking of different states' health agency policies intended to make more nutritional food available to young children whose families cannot afford it. All schools serve dry cornflakes with milk for breakfast. Informed by a model of comparative research that begins with naming and defining concepts, operationalizing them as variables, and hypothesizing relationships among dependent and independent variables before going to the field to test those hypotheses, you select your cases. In keeping with the positivist ontological and epistemological presuppositions underlying that model, you then assess the number of bowls of cereal consumed per day, by school type (public, church-sponsored) and/or location (city, town, rural), perhaps with additional independent variables thrown in for good measure. Your findings show some correlation between cornflake consumption and, say, children's height, weight, and grades, varying by type of school and rural-urban location, leading you to recommend particular courses of action intended to improve protein intake.

But your findings also show that in one state, school children eat very little of the cereal by contrast with other states' schools, thereby drastically lowering their nutritional intake. You assess your variables to see what might explain this finding, which remains puzzling to you – until a colleague whose research is informed by interpretive methodological presuppositions asks if you have taken the temperature of the milk into account. Having lived in that 'non-conforming' state for some time, in both city and village, as a participant-observer conducting a different research project, your colleague knows something of its eating culture and explains that local residents think it entirely inappropriate, not to say unhealthy and perhaps even unlucky, to begin the day by eating

131

cold food. Furthermore, warm milk turned the cornflakes mushy, much like wet paper, a taste deemed unpalatable. Not having your interpretive colleague's local knowledge, developed from living among and interacting with parents, teachers, children, and community elders and other leaders, you were unable to explain the finding, an explanation that to your colleague, with first-hand background in the setting, seemed self-evident.[1]

A set of increasingly widely known and cited 'rules' for the comparative analysis of social and political phenomena have been elaborated and codified in the last decade (for example, Mahoney and Goertz, 2004; George and Bennett, 2005; Gerring, 2007), as the chapters in this book also show. These methods rules have been articulated by scholars working within positivist traditions of social science – those that presuppose a realist social world and the possibility of knowing that world from a position external to the social realities being studied. To decide, for example, on a selection of cases that are most similar or most different, one needs to stipulate several things in advance of beginning field research, from that external position: (i) the concepts that are most relevant to what one wants to study; (ii) what they mean; (iii) how they are related; and (iv) how one will test those hypothesized relationships.

These points of departure pose difficulties, however, for researchers working within an interpretive scientific tradition that presupposes other methodological priors. This tradition presumes collectively, inter-subjectively constituted social 'realities' and the possibility of their know-ability only 'from within', from the contextualized perspective of the situational actor – the situated participant, the researcher herself, or, as in most research situations, both, as research-related knowledge develops as they interact. Consequently, research rests on requisite flexibility in research processes and unfolding learning over its course (Schwartz-Shea and Yanow, 2012). For such researchers, central to analysis is allowing the meanings of the key concepts and, often, the concepts themselves to emerge from the field, as discussed below.[2]

Such approaches to policy studies have become known as 'interpretive policy analysis' (Jennings, 1983, 1987; Healy, 1986; Torgerson, 1986; Yanow, 2000, 2007; Fischer, 2003). A broadly postpositivist perspective that places meanings – values, beliefs, and feelings or sentiments – at the centre of inquiry, this approach developed in the context of the wider interpretive turn taking place across the social sciences (see, for example, Geertz, 1973; Rabinow and Sullivan, 1979/1985; Hiley et al., 1991). Moving beyond a debate over analytic techniques, interpretive policy scholars call for a reconsideration of the aims of public policy studies

in favour of meaning-focused understandings of social and political realities resting primarily on word-based modes of analysis (see also Verloo, 2007; Bacchi, 2009). As Homi Bhabha explained with respect to the cornflakes, '... you have to know something about ... a place and its cultural rituals' in order to understand what is meaningful to situational actors – to know, that is, why some might prefer 'the sturdier Indian corn flake' with warm milk, even though its taste is perhaps inferior to 'Mr. Kellogg's corn flakes', or might even pass up the cornflakes altogether rather than eating them with cold milk (quoted in Schwartz-Shea and Yanow, 2012: p. 45).

This is not the place to rehearse methodological or philosophical critiques of positivist social science, which have been expounded elsewhere at length (for example, Hawkesworth, 1988, 2013/2006; Chabal and Daloz, 2006; Jackson, 2011; Yanow and Schwartz-Shea, 2013/2006a).[3] Instead, this chapter traces the origins of interpretive policy analysis and the methodological grounding for its approach and then turns to the question of what it might mean to do comparative research from an interpretive perspective.

Interpretive policy analysis and its methodological presuppositions

Central to interpretivist methodologies is the role that meanings – values, beliefs, and feelings (sentiments) – play in the understanding of social realities. Such an approach argues in favour of thickly contextualized renderings of social realities and of recognizing the inescapable subjectivity of the researcher as well as of the researched, along with the *inter*-subjective making of situated meaning. As one of the main evidentiary sources of policy meanings and their communication is language, interpretive policy analyses have largely engaged discourses, both written and oral, as well as non-verbal and other forms of language-based theories and methods; but evidentiary sources also include acts and material artefacts, and they are also included when they figure in the specific policy under consideration.

Interpretive policy analysis emerged in the United States in the late 1970s and early 1980s out of perceived limitations of the modes of analysis established at the beginning of the policy studies 'movement' of the late 1960s and early 1970s (seen in the founding of new postgraduate degrees and key journals, such as *Policy Sciences* in 1970 and *Policy Studies Journal* in 1972). Their evaluative tools – in particular, various forms of cost-benefit analysis, such as Policy-Planning-Budgeting System (PPBS) – rested on

the assumption that social values, quite aside from whether they were translatable into assessable measures, could be separated out from the realm of facts, which were, it was assumed, capable of being easily ascertained. But as the conceptual ground shifted concerning the possibility of disengaging values from facts (see, for example, Rein, 1976; Hawkesworth, 1988), some policy analysts, both academic and practice-based, began to see that these tools did not always work well for assessing certain features central to policy enactments and related practices. For example, city councils in Cambridge, Somerville, Ann Arbor, Santa Cruz, Hayward, Oakland, Berkeley, and other US cities adopted policies proclaiming themselves 'nuclear-free zones' and prohibiting the trans-shipment of nuclear waste across their borders. Nuclear material in the United States, however, is commonly transported over federally funded highways; and federal law, which regulates traffic on those roadways, takes precedence over local law, rendering those local policies unimplementable and – from the 'fact-value separation' perspective – 'meaningless'. Policy research or evaluation seeking to assess these legislative acts by establishing the facts of the case in order to determine the feasibility of the policy, whether in cost-benefit or other terms – on the assumption that values can be separated from facts and, effectively, isolated outside the realm of the study – would be of little use here as it would miss the point that these policies were all about values. To explain this further, I need first to bring in another issue.

A second source of dissatisfaction with a policy analysis built on positivist presuppositions emerged from tensions between empirical research into implementation processes, in particular, and the theoretically derived understanding of authority embedded in Weberian bureaucracy theory. This theorizing routinizes the power dimensions of policy implementation, rendering them either, for all intents and purposes, invisible or, when undeniable, as aberrations violating bureaucratic organizational principles. The model of the policy process that emerged, based on this separation of politics and power from organizational action, was instrumental-rational in character. It treated the policy process as a set of stages in a linear, assembly-line fashion marked by a top-down decision-making authority, what Shore and Wright (2011) call 'authoritative instrumentalism'. This thinking typically assumed that legislative intent is (or should be) capable of being made clear and known; that language itself is (or is capable of being made) transparent (with respect to its referent) and unambiguous; and that the policy process (from policy formulation through implementation) is exclusively rational and instrumental in its orientation towards stipulated goals. In this view, for governmental bodies to legislate policies that are incapable of being

implemented is irrational. In *Dance of Legislation*, Eric Redman presented such power dimensions, empirically observed and experienced in legislative processes, as an explicit part of policy making. His reflection, nearly 30 years after its publication, on readers' responses to his descriptions captures some of what from today's vantage point seems the same political, organizational, and conceptual innocence and naïve belief in Weberian theorizing characteristic of theoretical assumptions about policy implementation at the time:

> What strikes many readers of this book most forcefully was something I hadn't expected: surprise that bills do not advance strictly on their merits, that complex calculations of self-interest – perhaps having nothing to do with the merits, in fact – can be decisive in influencing a chairman or chief counsel or sponsor or staffer to aid this bill and not that, to choose one and drop others. This apparent revelation, the seeming arbitrariness of it all, provoked fascination and revulsion, sometimes in the same reader.
>
> (Redman, 2001/1973: p. 304)

The challenge that empirically grounded research and theorizing in policy implementation studies (for example, Pressman and Wildavsky, 1984/1973; Fox, 1990; Yanow, 1990) posed to the top-down, instrumental-rational model of policy making and implementation was joined by field-based studies of the work practices of implementers in various settings, including street-level bureaucrats (Prottas, 1979; Weatherley, 1979; Lipsky, 1980; for second-generation studies see Maynard-Moody and Musheno, 2003; Dubois, 2010/1999; see also Stein, 2004). These insights led Lipsky (1978) to argue – in a major critique of Weberian bureaucracy theory as applied to public policy processes – that the understanding of implementation needed to be inverted. As experienced and observed rather than as theorized absent empirical input, policies that were supposed to be being implemented in a-political administrative fashion were actually subject to local interpretation at the hands of street-level bureaucrats (due to particular constraints designed into bureaucratic organizational structures). These bureaucrats were then being understood by clients as themselves *making* governmental policy through their acts. The whole conceptual apparatus, in other words, needed to be re-thought, including with respect to what bureaucracy and other organizational theorists had argued was and should be a-political administrative practices.

Let's return to the nuclear waste shipment case. An evaluation of those cities' policies that insists on the separation of values from facts *and* that

rests on assessing their instrumentally rational goal-attainment would miss something central to them. We know that policy making, at whatever level of government, takes time and effort. Why would so many cities go to such lengths to behave irrationally, adopting policies that were irrational? Here is where an interpretive perspective is useful: it argues that policies, in general, are not only instrumental-rational acts, but are also expressive of human meaning. In this case, an interpretive analysis would focus on the meaning-ful character of these 'nuclear-free zone' policies as expressions of the values, beliefs, and feelings (or sentiments) of residents (including policy-makers) of these cities, as embodied in particular political identities; it would seek to articulate what those values, beliefs, and/or feelings are, from the perspective of those residents (by conversing with them, surely, but also by reading the documentary record of legislative debates and local newspaper coverage, perhaps also taking up residence among them and/or otherwise 'hanging out' on site in order to observe those values, beliefs, and/or feelings in everyday practices, so to speak); and it would explore who the audience(s) was (were) to whom those residents, through the legislated policies, sought to communicate those identities, beliefs, values, and feelings – in short, those meanings.[4] A meaning-focused approach to policy acts unsettles, then, the assumption built in to positivist modes of policy analysis that instrumental rationality is the sole, and solely legitimate, *raison d'être* of public policies. Interpretive policy analysis shifts the analytic focus in policy studies to meaning-making – its expression as well as its communication – seeing that policies and policy processes may also be avenues or vehicles for human expressiveness (of identity, of meaning). From an interpretive point of view, public policies can be understood as embodying and expressing the stories each polity tells itself and other publics, near and far, about its identity.

To be clear: the expressive dimensions of policies are rarely explicitly and intentionally crafted. There is little sitting down (in city council meetings, for instance) and saying, 'Let's figure out what is meaningful to us, what we want to express and how, so that we can adopt a "symbolic" policy – one that communicates values, beliefs, and/or feelings through their representations in language, objects, and/or acts'. All human action is potentially symbolic in that it is representative of embedded, often tacitly known, values, beliefs, and feelings. It is rare to have policies that are passed, explicitly and intentionally, for symbolic communication alone; that is, rational-instrumental action can also be, and at times is, symbolic as well. Interpretive analysis makes room at the scientific table for the expression and communication of meaning.

These critiques joined a third which had begun to develop at around the same time and from which interpretive policy analysis derives its name, the interpretive turn across the social sciences more broadly. This third set, critiquing (logical and neo-) positivist thought, drew on a range of ideas developing along parallel lines, some in other branches of political science, some in anthropology and sociology, others in philosophy, psychology, economics, and literary studies (for example, Edelman, 1964, 1971, 1977; Taylor, 1971; Geertz, 1973; Fay, 1975; Bernstein, 1976, 1983; Rabinow and Sullivan, 1979/1985; Gusfield, 1981; Fish, 1983; Polkinghorne, 1983, 1988). These included phenomenological and hermeneutic philosophies, joined by critical theory's focus on power; attention to symbols and their meanings within symbolic-cultural anthropology, semiotics, and literary studies; and pragmatism, ethnomethodology, and symbolic interaction's everyday action-meaning links. Interpretive policy analysis has incorporated additional elements from various other 'turns' that became central to social scientific thinking in the latter part of the twentieth century: linguistic, historical, metaphoric, practice, pragmatist, and so forth (see, for example, Fraser, 1995; McDonald, 1996; Lorenz, 1998; Schatzki et al., 2001; S. K. White, 2004). Today, one would also add science studies (the sociology of scientific knowledge), activity theory, and actor-network theory to the mix of linked turns and influences (see, for example, Brandwein, 2000, 2013/2006; Büger and Gadinger, 2007; Latour, 2008; Büger, 2012a; see also Miettinen et al., 2009).

'Interpretation', in this account, takes from hermeneutics mainly its focus on epistemic (or interpretive or discourse) communities, their intersubjective meaning-making (of texts and 'text-analogues'; see below) and potentially multiple, conflicting interpretations; the recursiveness of the hermeneutic circle; and indeterminacies of meaning (see, for example, Dryzek, 1982). Joined with phenomenology's insistence on the role of prior knowledge, including that generated through lived experience (acts, practices), in shaping meaning-making/interpretation, these ideas have proved generative for the understanding of public policies, their processes and practices, from affirmative action to whaling. In response to Habermasian theorizing and theoretical developments in other fields, a significant segment of interpretive policy analysis took a discursive, dialogical turn (for example, Fox, 1990; Fischer and Forester, 1993; Hajer and Wagenaar, 2003) which counters the denial of agency to those 'targets' on the 'receiving ends' of policies (see, for example, Sapolsky, 1972; Schneider and Ingram, 1993; by contrast, see Hacking, 1999). That move has, for several theorists, re-linked policy analysis

to forms of governance that are more participative and 'democratic', especially in their discursive focuses (Dryzek, 1990; Schneider and Ingram, 1997; Fischer, 2003, 2009).

In these several approaches, policy enactments (quite aside from their legislative record) may be usefully viewed as 'texts' (following Taylor's, 1971, argument about acts as 'text analogues'; see also Ricoeur, 1971), with implementers, clients, potential clients, and other policy-relevant publics, near and far, as 'readers' making sense of these 'texts' (Yanow, 1993). Many interpretive policy analysts, then, work to move beyond identification and description of communities of interpretation around specific policy issues and the understanding of what goes wrong (or right) in implementing them, to an exploration of communities of 'doing' and the specific practices that are entailed in the communication of policy meanings (Hajer, 1993; Yanow, 1995a; Freeman et al., 2011). In other words, at the same time that it 'tak[es] language seriously' (J. D. White, 1992) as one of the ways in which policy and implementing organizations' meanings are communicated, interpretive policy analysis also treats of acts and objects (material artefacts), such as programmes and built spaces, as communicative 'media'. (Some seek also to establish grounds for interventions to improve policy problems, akin to [partici-patory] action research; on that, see Greenwood and Levin, 2007.) In sum, interpretive policy analysts study various policy-relevant manifestations of the three broad categories of human artefacts – language, clearly, but also the acts (including non-verbal communication) and material objects that language references – which, through symbolic representation, give expression to their creators' meanings.

Interpretive policy analysis potentially draws, then, as much on participant-observer ethnography (useful in the study of acts and the material world) as it does on textual and other language-focused methods (such as discourse, metaphor/metonymy, rhetorical, and category analyses; for classic and contemporary examples of the former, see, for example, Selznick, 1949; Kaufman, 1960; Blau, 1963/1953; Crozier, 1964; Yanow, 1996; Soss, 2000, 2005, 2013/2006; Stein, 2004; Walsh, 2004, 2007; Dubois, 2010/1999; the chapters in Schatz, 2009 and Shore et al., 2011; and the essays in Symposium, 2009). In their various approaches to the expression and communication of what is situationally meaning-ful, interpretive policy analyses potentially take into account the local knowledge of those for whom policies have been designed and legislated, not necessarily restricting themselves to that of policy-makers and implementers (in both instances, depending on the research question). This often includes essaying to make what is

known tacitly, in Polanyi's sense (1966), more explicit. Interpretive policy analysis asks not only *what* a policy means – a context-specific question about a specific policy – but also *how* policies mean – questions about the processes by which meanings are learned, known, and communicated (Yanow, 1993, 1996; see also Torgerson, 1985). Borrowing a term from recent developments in cognitive linguistics, we might say that this focus on 'how' entails a *multimodal* form of analysis (Müller, 2008), looking at various sources and genres of evidence and corresponding analytic modes, something further theorized as the 'mapping' for exposure and intertextuality that characterizes interpretive methodologies and methods more broadly (Schwartz-Shea and Yanow, 2012). In addition, interpretive forms of policy analysis have shifted attention from the search for (and belief in the promise of finding) one correct policy formulation (correct in its definition of the policy problem, a framing that entails the seeds for problem resolution) to focus instead on the multiplicities of interpretation across policy-relevant groups. This includes perceiving the possibility that conflicts among such groups may reflect epistemological differences and not simply contests over facts. What is perceived and accepted as a relevant 'fact' is often part of the contestation, as Rein and Schon (1977) argued in respect of policy framing (see also Bacchi, 2009). Language, objects, and acts that are symbolic – that is, which represent unarticulated meaning(s) (values, beliefs, and/or feelings/sentiments) – enable multiplicities of (possible) meaning-making, a demarcation among communities of interpretation and of practice, and, hence, conflict (see, for example, Maynard-Moody and Stull, 1987).

To take one example, some years ago in a Jewish neighbourhood in Jerusalem, young, professional couples with newfound disposable income acquired pedigreed dogs as pets, which many of their neighbours, older couples who had survived the Nazi concentration camps, began to poison. Seeking to resolve the ensuing unrest and restore neighbourhood peace, a community organizer (possessed of what we might call an interpretive analytic bent) convened a meeting at which she asked one member of each group to talk about what the dogs meant to them. The representative of the young couples stood up and explained that in a socioeconomic context in which affording pets had been a luxury until very recently, owning a pedigreed dog had become a visible marker of their newly acquired middle-class status. One of the survivors then stood and explained that the 'German shepherds' these young couples had chosen as their pets were the very breed used by the Nazis during World War II, attacking relatives, friends, and even themselves, and that

seeing them around the neighbourhood brought back bitter memories. With each of the two parties to the conflict understanding the meaning of the dog for the other, a resolution was achieved: the young couples exchanged the German shepherds for another breed; the older residents stopped putting out poison (Leah Shinan-Shamir, personal communication, ca. 1979).

Here we see several characteristics of interpretive analysis: attending to meaning and its expression and communication through various modes of symbolic representation (objects and other elements in the non-human, material world; acts; language); exploring the potential for symbols to accommodate multiple meanings; according legitimacy to 'local' knowledge (that 'locality' shifting along with the research question); taking language and the cognition it voices seriously; and in practice- or intervention-oriented research, seeking to enable learning for purposes of enhanced communication and ensuing action. Much of the work on framing in interpretive policy analysis, discussed further below, explores the sorts of miscommunications that can arise from multiple interpretations of the same symbol (for example, Linder, 1995; Swaffield, 1998; on framing theory in policy analysis, see Rein and Schon, 1977; Schön and Rein, 1994; van Hulst and Yanow, 2014).

Recognizing the agency of those conceptualized as policy 'targets' and, perhaps even more importantly, treating their local knowledge as itself an important kind of expertise repositions the expertise of both practicing policy analysts and policy researchers, from purely subject-matter knowledge to knowledge of inquiry processes. In this fashion, the practice of interpretive policy analysis intertwines its conceptual approach with a set of methodological concerns and attendant methods that themselves focus on local knowledge. These generate data through the close reading (literal and/or figurative) of policy-relevant texts and other kinds of physical artefacts, conversational talk and formal interviewing, and participant–observer ethnography. To analyze those data, interpretive policy analysts draw on meaning-focused methods appropriate to the character of the data (for a list of such methods, see Yanow and Schwartz-Shea, 2013/2006a: p. xxiii; for specific treatment of several, see Yanow, 2000).

Data in the form of *language* predominate in interpretive policy analysis, in part due to the centrality to policy studies of documentary sources and, for interacting with legislators and others, of interviewing. Researchers recognize the inherent ambiguity of language – indeed, ambiguity in policy matters is often purposeful – and the ensuing multiple possible meaning-making of the same term or phrase. One stream of

language-focused research investigates the work of metaphors in policy language, much of it building on theories from cognitive linguistics (Lakoff and Johnson, 1980; Lakoff, 1993; Müller, 2008; in policy and political contexts, Edelman, 1977; Schon, 1979; Miller, 1985; Carver and Pikalo, 2008; Yanow, 2012/1992; Cienki and Yanow, 2013). Other work looks at categories, story-telling or other forms of narrative (Lakoff, 1987; Forester, 1993; Kaplan, 1993; Roe, 1994; Yanow, 1995b, 1999, 2003; Keeler, 2007; van Hulst, 2008; Rasmussen, 2011; Yanow and van der Haar, 2012). Framing is central to interpretive policy analysis, as noted above, building on the work of Rein and Schon (1977; Schön and Rein, 1994), whose theorizing pointed to the extent to which 'intractable policy controversies' are often that way not because of failures in policy design, but because of the particular way that the policy issue itself has been framed (Linder, 1995; Swaffield, 1998; Abolafia, 2004; Schmidt, 2013/2006). More recently, interpretive policy analysts have taken up discourse theories of various sorts (Howarth, 2000; Epstein, 2008).

To analyze policy-related *acts*, such as the act of choosing or declining to regulate electro-magnetic frequency emissions (Linder, 1995), beyond analyzing whatever legislative or other language might be in play, researchers might draw on ethnographic or participant–observer analysis of the various groups involved (see, for example, Schatz, 2009). Furthermore, language and acts also often invoke or use *objects* in the material world. Examples might include the way photographs depict welfare recipients in the context of welfare policy reform (Schram, 2002), the specific design of policy-relevant spaces, or the meaning of 'home', whether ownership or occupancy, in a particular housing policy that enables either purchase or rental. For such data, analysis might focus on the ways in which programmatic activities or built spaces communicate policy and/or wider societal meaning(s) and which meanings are being communicated and to what audiences, near and far (Goodsell, 1988, 1993; Yanow, 1995b, 2013/2006). Interpretive policy analysts are also turning to ethnographic observation that, adapted from science studies and actor-network theory (Latour and Woolgar, 1988; Nicolini, 2009), follows the policy, key policy-relevant material artefacts or actors.

These and other analytic methods are useful in trying to elicit understandings of what specific policies might mean to various issue-relevant publics as well as in exploring how those meanings are developed, communicated, and (potentially) variously understood (Yanow, 2000, 2007). Through them, analysts seek to 'map the architecture' (Pal, 1995) of

policy arguments (for empirical examples, see Linder, 1995; Swaffield, 1998; Epstein, 2008). Distinguishing among the three categories of data types is useful heuristically even though they are often not, in practice, fully distinct: language, acts, and objects are intertwined and mutually implicating, and whether one designates a bit of policy evidence as belonging in one rather than another category may at times make sense only from the perspective of the analysis one is trying to advance.

In these several analytic treatments, the notion of policy, whether legislative document or state intention, as a single, authored text is implicitly replaced by the 'constructed' texts of multiple 'readings' at the hands of various policy-relevant publics. The notion of a singular (that is, collective) legislative 'author' is joined by multiple discourse communities in the form of collective 'readers'. All analyses emphasize the context-specificity of meaning. They are specific to events and times – the 'what' of a policy – and hew closely to the meanings made by policy-relevant actors, although an analysis may be, and often is, more broadly contextualized, whether by reference to multiple evidentiary sources in a comparative vein or to the context of some theoretical literature to which the research question and analysis speak. Generalization typically bears on the 'how', and it tends to be more a generalization concerning meaning-making *processes*: legislators, implementers, clients, and other publics make policy meanings in thus and such ways, *or* discourses play out in this case in thus and such ways; some of these interpretations or discourses explain how and why conflicts took place (for example, in the dog-poisoning case); and here are ways in which analysts might make sense of those meanings and conflicts and, perhaps, even suggest resolutions to the conflicts or improvements in practices. In showing, for example, how agency policy-makers and founders expected the Israel Corporation of Community Centres' buildings and programmes to 'perform' or enact the values to which residents of the 'impoverished' towns and neighbourhoods in which they were set up could (and should, in the founders' views) aspire, by contrast with many residents' other interpretations of those centres, I could explain difficulties (at the time) in implementing policy programmes (Yanow, 1996). The analysis showed the role that the communication and interpretation of policy and organizational meanings played in policy settings, a meaning-making process that could be generalized beyond that specific setting (along with a methodological orientation, a set of methods, and some potentially key analytic concepts).[5]

One final methodological point. Although phenomenology, in particular, has been critiqued, especially by critical theorists, for being so

involved with the individual 'self' as to neglect power, this criticism, although perhaps holding at the philosophical level, seems not to obtain when phenomenological inquiry is directed towards policy matters. In such applications, analysts cannot help but include power dimensions, including of organizations and other institutions, especially when they consider voices that have been silent, by choice, or silenced, by forces beyond them. This orientation towards power and the political can be seen in the discourse analysis and other empirical research examples described and cited throughout this chapter.

Implications for designing interpretive comparative research

There are so many different ways of doing interpretive research that no single research design serves them all.[6] Consider, for instance, the research design implications of positivist-informed approaches to comparative research, which stipulate the definition of concepts as a starting point. Interpretive research has no such singular starting point. Instead, a researcher enters the hermeneutic-interpretive circle-spiral at any starting point, with whatever (prior) knowledge she has at that moment. Rather than a deductive or inductive rationale, interpretive research follows an abductive logic of inquiry: it begins with a puzzle, a surprise or a tension, typically arising from the juxtaposition of expectations – themselves deriving from *a priori* knowledge, whether theoretical or experiential – with field observations, experiences, and/or readings (for extended discussion see Schwartz-Shea and Yanow, 2012: ch. 2). In light of this lack of singularity, and because research is a practice whose 'rules' can only be learned and known in the practicing (as Büger, 2012b, makes clear), a discussion of interpretive research design can only outline some key concerns, making them clear by contrast with the tradition outlined in most methods textbooks and referring readers to other sources.

For a methodology that puts situated meaning front and centre, a method focused on deriving universally generalizable concepts is problematic, at best. Comparison, from an interpretive perspective, cannot begin by presuming equivalences between or among polities. (Whether these are equivalences of similarity or difference is immaterial: both of these measures rest on posited identities, a point Miller, 1985, makes clear with respect to exchange and other metaphoric relationships that have clear analogies in comparative research.) Comparison must work 'from the [researcher's] ability to make sense of the singularities of each system, rather than from the capacity either to slot them into pre-determined boxes or to place them on a continuum' (Chabal and Daloz, 2006: p. 63).

Interpretive research refuses to lose the local or historical specificities from which concepts emerge.[7] I return to the matter of generalizability below.

This leads to one of the central hallmarks of interpretive research designs, which has crucial implications for planning comparative research: the relative openness of the research process, with the consequent requisite flexibility that is built into the design and execution of such research. Researchers need to be able to respond to field situations. If an archived document points to another, hitherto unknown archive, the researcher needs 'research design flexibility' to follow that lead. If an organizational, community, or other member invites the researcher to join in a particular activity, he needs 'research design flexibility' to accept (or decline) the invitation. It is not possible to plan for such occurrences ahead of time. What one plans for, indeed, is judgement-making on the spot. Zirakzadeh (2009) could not have anticipated being shot at by Spanish police, nor did he anticipate and plan on becoming a local celebrity in the bar that evening – resulting in leads to all manner of 'evidentiary sources' – or on thereby becoming persona non grata among other cadres of the Basque separatists he was studying. Such openness and flexibility are neither an artefact of poor research design nor a failing of the researcher. Interpretive research is, and has to be, open to the researcher's on-the-spot judgements – whether in observational or archival fields.

Their privileging of 'local' or situated meaning leads interpretive researchers, including policy analysts, to the methodological insistence on 'allowing concepts to emerge from the field'. This statement is used to designate a counter-position to one that begins by defining concepts a priori, operationalizing them as variables, and hypothesizing relationships among those variables, in order to test those hypotheses in the field, in a deductive logic of inquiry, against social 'realities'. Comparative research in a positivist mode rests on this kind of *a priori* assumption of knowledge: in order to ascertain which cases will be most similar or most different, or any other form of equivalence, the researcher needs to know enough beforehand in order to stipulate not only what the independent and dependent variables will be, but also which ones are most likely to serve as points for comparison. This 'front-loads' research learning in the design process.

Interpretive researchers, by contrast, assume that learning will continue throughout their research. Rather than being front-loaded, taking place before and as the research design is being developed, learning is expected to unfold throughout the research process, in encounters

during fieldwork with situated members, activities, documents, and so forth and continuing through deskwork (analysis) and textwork (writing) phases (Schwartz-Shea and Yanow, 2012: ch. 4). What is being learned are the specific, local meanings of the event(s), act(s), interactions, and so on, on which the research focuses and their relationship to its theoretical frame. In this sense, comparative research in an interpretive mode could well be understood as crossing semantic borders, regardless of whether or not it crosses national ones (Rutgers, 2004: p. 154). Researchers expect to learn, for instance, how situated actors see the matter under investigation and what they consider meaningful in it. This, too, means a limit to what can be presumed and stipulated *a priori*. Moreover, through research designs and their implementation in observational or archival fields, as well as later during deskwork and textwork, interpretive researchers essay to avoid the 'rush to diagnosis' that would result in the 'premature closure' of inquiry. Predetermining the relevant axes of comparison would be precisely such a foreclosure in a study that is interested more in 'local knowledge' and concept definition-in-use than in the researcher's own *a priori* concept definition.

'Comparison' takes on a particular meaning in this methodological context. Due to its explicit acknowledgement and incorporation of the researcher's prior knowledge and expectations, interpretive research is innately and inherently always, constantly comparative, in a couple of respects. The first of these – the comparison of researcher expectations with field 'realities' – has typically not, however, been incorporated explicitly in research designs. The initial point of comparison – the researcher's *a priori* knowledge and own political-cultural context – commonly remains unspoken, often being known and appreciated only tacitly. To the extent that interpretive methodologies increasingly call for explicit reflection on the researcher's 'positionality' with respect to epistemological matters and knowledge claims – whether positioned 'geographically' (for example, being on the processing floor or on the quality-control catwalk; Pachirat, 2011), demographically (for example, education, class, sex, birthplace, and so on; Shehata, 2013/2006), or theoretically (for example, drawing on a social movement frame; Zirakzadeh, 2009) – we might expect to see this initial point of comparison made increasingly explicit. Abductive logic potentially makes part of this explicit by positioning puzzles or surprises front and centre in a research project, thereby encouraging researchers to account for the sources of their puzzlement. As abduction is more widely taken up to explain interpretive methodologies, we should see more explicit attention to the constant comparison embedded in such juxtapositions.

The second aspect of interpretive research's constant comparison is internal to the 'case' being studied. Interpretive research projects typically take up only one case at a time because of the extensive immersion in a setting required to fathom local meaning-making and because of the ensuing degree of detail required to account contextually for that setting and what is meaningful there (reflected in the mounds of data in field notes recording observational, interview, and/or documentary detail, and variances in field procedures from the research design, for example, due to the unexpected absence or unanticipated entry of key actors). Although the rationale for single cases is self-evident to interpretive researchers, it often becomes a point of difficulty when they are asked by those unfamiliar with this logic of inquiry to justify their methods.

From other methodological perspectives, such research is commonly designated 'single n' research: the 'n', in language borrowed from experimental research design, designates the number of observations made in a research project; the greater the 'n' subjected to analysis, the more robust the knowledge claims derived from the research – that is, the more valid its general principles, as judged on the grounds of statistical inferences. That understanding, however, rests on a particular logic of scientific inquiry, one that is different from the logic of interpretive science; it also betrays a misunderstanding of the character of 'n' in interpretive research. Although the research may be conducted in a single setting, the 'n' entails far more than just a single 'observation', as that term originally meant in experimentation. Consider a relatively short-term field research project of six months' duration (by contrast with a more customary year-long immersion or even longer). In that time, a researcher may observe for 120 days (five days per week × four weeks per month × six months) and, say, 960 hours (at eight hours per day) – more, when conducting the more typical observational research project '24/7'. Either of these constitutes far more than a single observational 'n' – even if the research is being conducted on a single policy, state or other case setting. Or, considering observational moments, acts, interactions, and movements analyzed, casual conversations and/or interviews conducted, documents read and studied, the 'n' is also more than one – especially if it refers not to the sum total of interviews or time in talk, documents, and practices and objects engaged, but to the number of words, acts, and physical artefacts to be analyzed.

Comparison is ongoing across all of these elements, as the interpretive researcher 'maps' research settings for exposure to a variety of possible meanings/interpretations within them concerning the matter of interest. In organizational or community studies, where analysis often compares

different departments or neighbourhoods, such mapping for exposure could be horizontal or vertical across a bureaucratic organization or spread geographically within a community (or other polity). Such design searches for intertextuality (or its absence). This could be literal as when multiple documentary genres use the same or similar language across time periods or setting locations (or don't). Kathy Ferguson (2011) captured the comparative character of this research when voicing her exhilaration in the moment of discovery in archival research: 'Oh, look, there's another one!' In a more figurative sense, intertextuality is manifested when the researcher encounters 'cross-referencing' across data genres, for example, identical key words or phrases appearing across talk and print sources.

Comparison is inherent and constant, then, not only between the researcher's prior knowledge – of her home setting and its meanings and cultural practices, whether national or organizational; of her theoretical frame – and the setting under study and its events and circumstances, but also across multiple 'sites' and various sources and genres of data within the one setting, even when engaging 'only' a single case. The 'n' of observations in an experiment has been conflated, mistakenly, with the 'n' of settings in a field research project (Yanow et al., 2009), with consequent erroneous judgement concerning the scientific adequacy of a single-site study and of the character of comparison in such a study.

Rather than focusing on 'case selection,' then, interpretive research design expends considerable attention on questions of access, starting with choices of settings, actors, events, archives, and texts in which and among whom to pursue the research question (see Schwartz-Shea and Yanow, 2012: ch. 4), as well as the most appropriate timing for adequately studying that case and its full roster of 'n observations' (persons, activities, documents, and so on). Interpretive policy analysts can be as interested in understanding what is happening with policies 'on the ground', 'studying down', as they are in looking 'at the top', 'studying up' (as anthropologists put it; Nader, 1974/1969) where legislative decisions are framed and made; implementation studies can entail both. Participant-observer, ethnographic research is particularly useful in these as it generates a fine-grained evidentiary base that can support robust inferences. Shore has argued that policies require not so much studying up as studying across and every which way in a network sort of fashion. This is what following a policy and its relevant actors, objects, acts, and language promises, 'teasing out connections and observing how policies bring together individuals, discourses and institutions … and the new kinds of networks, relations and subjects this process creates' (2012: p. 172).

Following policy 'components' in these ways is a particular form of ethnographic method. Adapted into interpretive policy and other analyses from science studies – or, perhaps, a new way to talk about an older method used in early bureaucracy studies (for example, Kaufman, 1960) – it can lead analysts to trace the sites of agenda setting, decision making, and other loci of power and of silent and/or silenced voices without constraining the study within the borders of a specific physical setting. The policy itself is the site, not some geographically bounded entity, as is amply evident in, for instance, Epstein's (2008) discourse analytic study of whaling. Tracing how a policy issue might be framed at one moment and reframed at another can transcend both physical boundaries and those of time.

Comparative analysis of two or more cases may also unfold through sequential studies across the span of a research career. Schaffer's inquiry into the meaning(s) of 'democracy' when 'translated' from Western states to others is an example. In his first study, focused on Senegal (Schaffer, 1998), the comparison is rendered explicit in his research question. His subsequent research in the Philippines adds another locus of comparison to his theorizing concerning the meanings of 'democracy' as practiced in different places (Schaffer, Forthcoming).

The need for a comparative research design that stipulates *a priori* points of comparison appears to be driven by the desire to generalize findings across cases; and the need to generalize entails establishing causal relationships of a particular, mechanistic sort. All of these are conceptually and methodologically problematic from the perspective of interpretive research. Again, experimentation serves as the model for the assumptions that underlie generalizability: that the world being studied is or can be rendered stable (achieved through researcher controls over research design, choice of participants and control groups, and so on), and that both researchers and 'subjects' are interchangeable with other researchers and 'subjects'. Furthermore, that stability also presupposes cases that are discrete, such that interactions between them with respect to the focus of study will not contaminate findings.

By contrast, interpretive analyses assume a research setting that is dynamic and which cannot be subjected to researcher control, and a learning process in which knowledge, both researchers' and participants', is situated – specific to the setting in and from which it has been generated. This does not rule out the possibility that concepts have travelled – for example, that the meanings of 'democracy' in Senegal, the Philippines, and the United States might have influenced each other (see also Adcock, 2006, 2011). In fact, following conceptual travel would be of great

interest to an interpretive researcher. With respect to causality, where positivist comparative research is in search of mechanical causality – akin to a cue hitting a ball on a pool table – interpretive research is after constitutive causality: explanations that rest on the 'interweaving of codes [of meaning] in particular situations' rather than being 'logically derivable from the codes themselves' (Hammersley, 2008: p. 55; see Schwartz-Shea and Yanow, 2012: chs. 3, 6). As Rudolph and Rudolph put it in their interpretation of Weber's methodological position, 'an "understanding" (*verstehende*) interpretation of human events reaches to relationships and regularities but not to necessity; that is, it is governed by meaning, not by laws' (2008/1979: p. 173; quoted in the chapter epigraph).

For research done from a methodological perspective animated by a concern for situated meaning which intends to compare two (or more) cases in equivalent detail, the *a priori* designation of what is locally meaning-ful is a logical contradiction; it provides an inadequate design rationale. Instead, pursuing an abductive logic of inquiry, the interpretive comparative policy (or other) analyst would look during the research for additional settings relevant to the policy element being tracked which might shed further light on the initial 'surprise', showing further, even unanticipated, dimensions of the subject of study. The question is, where else might X – the event, issue, physical artefact, even concept of interest – be meaningful in key ways? Where else might the researcher find cases of this X which might expand the understanding of its range of meanings or rationales for action of particular sorts? In keeping with an abductive logic of inquiry, however, the comparative interpretive researcher is open to being surprised – even with the surprise that X is not at all operative in the new setting. One might even say that interpretive researchers assume that a different case setting will, *ipso facto*, generate different situated meanings (although the meaning-making processes in different settings might be similar). Whether the differences will be great (most different) or small (most similar) cannot be determined *a priori*.

Concluding thoughts

In sum, interpretive policy analysis is more of an approach than a stepwise method. An umbrella term from the perspective of analytic methods, it encompasses too many of these to lay out a general, step by step design (although such detail can be found in literature engaging specific methods, for example, semiotics, ethnomethodology, metaphor analysis, category analysis, and so on.). Because interpretive methodology insists

on researcher learning throughout a research project, requiring an openness and flexibility in its research design and allowing concepts to 'emerge from the field', there is also no step by step process for selecting cases; that activity is based on a different logic of inquiry. Moreover, these ontological and epistemological differences make it logically inconceivable to combine interpretive and positivist methodologies and their attendant methods in the same research question (although they may be used to address different questions within the same, broad research project; on this point, see Jackson 2011: 208-10, Schwartz-Shea and Yanow, 2012: ch. 8).

No science is independent of the setting in which it develops, and that includes the social, cultural, and political dimensions of that setting, as Adcock (2011, 2013/2006; Adcock and Vail, 2012) and Oren (2003) have amply demonstrated. The desire to stipulate ahead of time the steps that a researcher will undertake harks back to the matter of control that lies at the heart of science in its emergence, historically, against the back-drop of judgements made at the hands of monarchs and others in medi-eval European feudal and church hierarchies. The eighteenth-century Europe out of which positivist philosophies and 'science' – initially natural and social 'philosophy' – emerged was driving towards the 'enlightenment' project that set science against religion, later juxtapos-ing scientific 'testability' against religious 'tests' of faith. Natural and social philosophies and their subsequent scientific counterparts shifted the locus of judgement and meaning-making to individuals' powers of reason, unmediated by religion-supported authorities and liberated from metaphysical explanations.

Part of this development required the separation of facts from values. Control, however, did not disappear from the picture. It took up a central place in laboratory experimentation and research design, in which researchers control the cells in their Petri dishes, the mice in their mazes, and later, the human subjects in their pseudo-electric-chairs and mock prisons (see Yanow and Schwartz-Shea, 2008). This telescopic view of the development of 'science' makes clear that the routinization of scien-tific procedures, first in experimentation with its 'controls,' later in other forms, has been a long-term project intended to eliminate, more and more, the human hand (so to speak) from those procedures as well as to rationalize the power of the scientist through the authority of science.

In interpretive research, however, researchers have little control over the places they go to study or the 'occupants' of those places: an interpretive researcher cannot rearrange a research setting or dismiss a participant who is native to that setting, for instance, in the way that an experimentalist

can redesign a laboratory or eliminate a potential subject who does not fit the study's parameters (Schwartz-Shea and Yanow, 2012). What perhaps challenges positivist research most profoundly about interpretive research is the latter's willingness to 'break through the fourth wall,' in theatre-speak: through involving 'participants' and their knowledge more centrally in the shared work of knowledge-generation, thereby displacing researchers from centre stage, to call attention to the separation between researcher power and that of those on or among whom research has been, and is being, carried out and to dissolve that separation, accepting the limits of researcher control and knowledge. Whereas positivist-inflected or -influenced methods, in general, disappear 'the political' and 'the emotional' from science, interpretivist-inflected methods, on the whole, seek through these sorts of processes to make the existence and presence of power and emotions more transparent.

Reader-response theory (for example, Iser, 1989), which enables that breakdown of the fourth wall in adding readers' meaning-making to the question of where meaning comes from, lays the conceptual groundwork for seeing that it is, in fact, the whole project of interpretive methodologies and methods that is seeking to break through that wall, which occludes matters of power and control. The increasing 'rule-ification' of comparative analysis, through more different/more similar and other design tactics, seeks after a 'routinization of power' within comparative research. Much as Weber intended with respect to authority in bureaucratic organizational structures, rendering comparative method so systematic disappears the dance of power that is inherent in research.[8] As Robert Adcock notes (personal communication, 9 April 2011), 'political scientists do political work indirectly through the way they structure transatlantic comparisons to tell competing stories about how nations "lead" or "lag" one another'. The growing systematization of comparative method similarly achieves political work, in similarly indirect, non-explicit ways.

These observations speak to some of the political dimensions of science. Positivist methodologies and methods write the power issues out, routinizing them (in a Weberian sense) in the process of reducing the chaos and complexity of the human world to a scientific one subject to human controls. Interpretivism seeks to write that human element back in (Pachirat, 2013/2006; Yanow and Schwartz-Shea, 2013/2006c), even, in some cases, to the point of making sure that the messiness of the human world remains part of the science that seeks to understand and explain it (a point emphasized by Law, 2004). From this perspective, interpretive comparative research design and analysis will, and must,

perforce be carried out in radically different, but no less systematic and scientific ways from those prescribed by a comparative method informed by positivist presuppositions about social realities and their know-ability. It is a different systematicity, a different science – and difference need not be seen as inferior, nor, as Minow (1991) has argued in a policy context, need it be feared.

Notes

1 My thanks to Peri Schwartz-Shea for help in turning Homi Bhabha's story about eating cornflakes in India (quoted in Schwartz-Shea and Yanow, 2012: p. 45) into this imaginary comparative research design story.

2 This chapter rests on three nomenclature points. One, *methodological* refers to the ontological and epistemological presuppositions that inform specific *methods* (or 'tools'). Two, on the use of *interpretive* rather than *qualitative*, see Yanow (2003a), Yanow and Schwartz-Shea (2013/2006b); on other terminological and conceptual differences and their implications for research design, see Schwartz-Shea and Yanow (2012). Three, I use *field* in reference both to interviewing and observational research (with whatever degree of participation) and to archival research (considering not only documentary but also visual materials such as paintings; see, for example, Bellhouse, 2011).

3 Concerning the historical grounding of comparative analyses, Adcock's reading of the development of comparative method (2006, 2011) may serve further to extend these critiques. See, in particular, his reading of Mills' methods (Adcock 2009), commonly cited as the basis for comparative research.

4 There is nothing to suggest that expressive acts cannot also be instrumental (or vice versa). For example, a campaign that included hanging anti-nuclear signs on numerous homes, streets, schools, and offices across New Zealand influenced then-newly-elected Prime Minister David Lange's decision to change his stance on the matter, with a subsequent change in national policy (Raymond Nairn, personal communication, July 1996). Or, US policies establishing, defining, and using race-ethnic categories enable the polity and member groups to tell various kinds of identity stories (national, group, and origins; Yanow, 2003) at the same time that they enable very instrumental actions, as in the census and various affirmative action-related practices in hospitals, police work, and so on.

5 This discussion raises questions concerning interpretive comparative historical analysis which I do not have space to explore. See Linch (2011) on the importance of not treating the past 'as a laboratory of test cases, without regard for the cultural, linguistic and conceptual challenges to its scrutability'.

6 This and the next paragraphs draw on ideas that are more fully developed in Schwartz-Shea and Yanow (2012).

7 Interestingly, Woodrow Wilson was of the opinion, according to Adcock (2006: p. 128), that comparative method in US political science should retain the historical context of the subject of study. Then again, this precedes the separation of history and political science into independent disciplines.

8 An exchange with Robert Adcock helped me clarify this thought. On Weber's success in this effort (or lack thereof), see Rudolph and Rudolph (2008/1979).

Bibliography

Abolafia, Mitchel Y. (2004) 'Framing Moves: Interpretive Politics at the Federal Reserve', *Journal of Public Administration Research and Theory*, 14 (3), 349–70.

Adcock, Robert (2006) 'The Emigration of the "Comparative Method"', *European Political Science*, 5 (2), 124–36.

Adcock, Robert (2009) 'Mills' Methods Redux', Western Political Science Association Annual Conference, March, Vancouver.

Adcock, Robert (2011) 'Liberalism, Democracy, and the Transatlantic Gaze', Tulane University Presentation, 8 April.

Adcock, Robert (2013/2006) 'Generalization in Comparative and Historical Social Science', in Dvora Yanow and Peregrine Schwartz-Shea (eds) *Interpretation and Method*, 2nd edn (Armonk: M.E. Sharpe), 80–96.

Adcock, Robert and Vail, Mark (2012) 'Beyond Pluralism? Corporatism, Globalization, and the Dilemmas of Democratic Governance', in Mark Bevir (ed.) *Modern Pluralism: Anglo-American Debates Since 1880* (New York: Cambridge University Press), 214–46.

Bacchi, Carol Lee (2009) *Analysing Policy: What's the Problem Represented to Be?* (Frenchs Forest: Pearson Education).

Bellhouse, Mary L. (2011) 'Under the Eaves of the Louvre: Political Theory Research in American and European Art Museums, Art Collections, and Art Libraries', *Western Political Science Association Annual Conference*, 21–23 April, San Antonio.

Bernstein, Richard J. (1976) *The Restructuring of Social and Political Theory* (Philadelphia: University of Pennsylvania Press).

Bernstein, Richard J. (1983) *Beyond Objectivism and Relativism* (Philadelphia: University of Pennsylvania Press).

Blau, Peter (1963 [1953]) *The Dynamics of Bureaucracy* (University of Chicago Press).

Brandwein, Pamela (2000) 'Disciplinary Structures and "Winning" Arguments in Law and Courts Scholarship', *Law and Courts*, 10 (3), 11–19.

Brandwein, Pamela (2013/2006) 'Studying the Careers of Knowledge Claims: Bringing Science Studies to Legal Studies', in Dvora Yanow and Peregrine Schwartz-Shea (eds) *Interpretation and Method*, 2nd edn (Armonk: M.E. Sharpe), 284–99.

Büger, Christian (2012a) 'From Epistemology to Practice', *Journal of International Relations and Development*, 15 (1), 97–109.

Büger, Christian (2012b) 'Art, Methodology, and the Practice of Designing (Interpretive) Research', *Qualitative and Multi-Method Research*, 10 (2), 12–14.

Büger, Christian and Gadinger, Frank (2007) 'Reassembling and Dissecting: IR Practice from a Science Studies Perspective', *International Studies Perspective*, 8 (1), 90–110.

Carver, Terrell and Pikalo, Jernej (eds) (2008) *Political Language and Metaphor* (London: Routledge).

Chabal, Patrick and Daloz, Jean-Pascal (2006) *Culture Troubles: Politics and the Interpretation of Meaning* (University of Chicago Press).

Cienki, Alan and Yanow, Dvora (eds) (2013) 'Linguistic Approaches to Analysing Policy and the Political', Special Issue, *Journal of International Relations and Development*, 16 (2), 167–310.

Crozier, Michel (1964) *The Bureaucratic Phenomenon* (University of Chicago Press).

Dryzek, John S. (1982) 'Policy Analysis as a Hermeneutic Activity', *Policy Sciences*, 14 (4), 309–29.

Dryzek, John S. (1990) *Discursive Democracy* (Cambridge University Press).

Dubois, Vincent (2010/1999) *The Bureaucrat and the Poor* (London: Ashgate).

Edelman, Murray (1964) *The Symbolic Uses of Politics* (Urbana: University of Illinois Press).

Edelman, Murray (1971) *Politics as Symbolic Action* (Chicago: Markham).

Edelman, Murray (1977) *Political Language* (New York: Academic Press).

Epstein, Charlotte (2008) *Power of Words in International Relations* (Cambridge: MIT Press).

Fay, Brian (1975) *Social Theory and Political Practice* (Boston: George Allen & Unwin).

Ferguson, Kathy (2011) 'The Gift of the Archive', presentation, Western Political Science Association Annual Conference, 21–23 April, San Antonio.

Fischer, Frank (2003) *Reframing Public Policy* (New York: Oxford University Press).

Fischer, Frank (2009) *Democracy and Expertise: Reorienting Policy Inquiry* (New York: Oxford University Press).

Fischer, Frank and Forester, John (eds) (1993) *The Argumentative Turn in Policy Analysis and Planning* (Durham: Duke University Press).

Fish, Stanley (1983) *Is There a Text in This Class? The Authority of Interpretive Communities* (Cambridge: Harvard University Press).

Forester, John (1993) 'Learning from Practice Stories: The Priority of Practical Judgment', in Frank Fischer and John Forester (eds) *The Argumentative Turn in Policy Analysis and Planning* (Durham: Duke University Press), 186–209.

Fox, Charles J. (1990) 'Implementation Research', in Dennis J. Palumbo and Donald J. Calista (eds) *Implementation and the Policy Process* (New York: Greenwood), 199–212.

Fraser, Nancy (1995) 'Pragmatism, Feminism, and the Linguistic Turn', in Seyla Benhabib, Judith Butler, Drucilla Cornell, and Nancy Fraser (eds) *Feminist Contentions* (New York: Routledge), 157–72.

Freeman, Richard, Griggs, Steven, and Boaz, Annette (eds) (2011) 'The Practice of Policy-Making', Special Issue, *Evidence & Policy*, 7 (2), 127–227.

Geertz, Clifford (1973) *The Interpretation of Cultures* (New York: Basic Books).

George, Alexander and Bennett, Andrew (2005) *Case Study and Theory Development* (Cambridge: MIT Press).

Gerring, John (2007) *Case Study Research* (New York: Cambridge University Press).

Goodsell, Charles T. (1988) *The Social Meaning of Civic Space* (Lawrence: University Press of Kansas).

Goodsell, Charles T. (ed.) (1993) 'Architecture as a Setting for Governance', Special Issue, *Journal of Architectural and Planning Research*, 10 (4), 271–356.

Greenwood, Davydd J. and Levin, Morten (2007) *Introduction to Action Research*, 2nd edn (Thousand Oaks: Sage).

Gusfield, Joseph R. (1981) *The Culture of Public Problems* (University of Chicago Press).

Hacking, Ian (1999) *The Social Construction of What?* (Cambridge: Harvard University Press).

Hajer, Maarten (1993) 'Discourse Coalitions and the Institutionalization of Practice', in Frank Fischer and John Forester (eds) *The Argumentative Turn in Policy Analysis and Planning* (Durham: Duke University Press), 43–76.

Hajer, Maarten A. and Wagenaar, Hendrik (eds) (2003) *Deliberative Policy Analysis* (New York: Cambridge University Press).

Hammersley, Martyn (2008) *Questioning Qualitative Inquiry* (Thousand Oaks: Sage).

Hawkesworth, M. E. (1988) *Theoretical Issues in Policy Analysis* (Albany: SUNY Press).

Hawkesworth, Mary E. (2013/2006) 'Contending Conceptions of Science and Politics', in Dvora Yanow and Peregrine Schwartz-Shea (eds) *Interpretation and Method*, 2nd edn (Armonk: M.E. Sharpe), 27–49.

Healy, Paul (1986) 'Interpretive Policy Inquiry', *Policy Sciences*, 19 (4), 381–96.

Hiley, David R., Bohman, James F., and Shusterman, Richard (eds) (1991) *The Interpretive Turn* (Ithaca: Cornell University Press).

Howarth, David J. (2000) *Discourse* (Buckingham: Open University Press).

Iser, Wolfgang (1989) *Prospecting: From Reader Response to Literary Anthropology* (Baltimore: Johns Hopkins University Press).

Jackson, Patrick Thaddeus (2011) *The Conduct of Inquiry in International Relations* (New York: Routledge).

Jennings, Bruce (1983) 'Interpretive Social Science and Policy Analysis', in Daniel Callahan and Bruce Jennings (eds) *Ethics, the Social Sciences, and Policy Analysis* (New York: Plenum), 3–35.

Jennings, Bruce (1987) 'Interpretation and the Practice of Policy Analysis', in Frank Fischer and John Forester (eds) *Confronting Values in Policy Analysis* (Newbury Park: Sage), 128–52.

Kaplan, Thomas J. (1993) 'Reading Policy Narratives', in Frank Fischer and John Forester (eds) *The Argumentative Turn in Policy Analysis and Planning* (Durham: Duke University Press), 167–85.

Kaufman, Herbert (1960) *The Forest Ranger* (Baltimore: Published for Resources for the Future by Johns Hopkins Press).

Keeler, Rebecca (2007) 'Analysis of Logic: Categories of People in U.S. HIV/AIDS Policy', *Administration & Society*, 39 (5), 612–30.

Lakoff, George (1987) *Women, Fire, and Dangerous Things* (University of Chicago Press).

Lakoff, George (1993) 'The Contemporary Theory of Metaphor', in Andrew Ortony (ed.) *Metaphor and Thought*, 2nd edn (New York: Cambridge University Press), 202–51.

Lakoff, George and Johnson, Mark (1980) *Metaphors We Live By* (University of Chicago Press).

Latour, Bruno (2008) *Reassembling the Social: An Introduction to Actor–Network Theory* (New York: Oxford University Press).

Latour, Bruno and Woolgar, Steve (1988) *Laboratory Life*, 2nd edn (Princeton University Press).

Law, John (2004) *After Method* (London: Routledge).

Linch, Amy (2011) 'Interpreting Across Time: The Use of History in Comparative Democratization', Western Political Science Association Annual Meeting, 21–23 April, San Antonio.

Linder, Stephen H. (1995) 'Contending Discourses in the Electromagnetic Fields Controversy', *Policy Sciences*, 28 (2), 209–30.

Lipsky, Michael (1978) 'Standing the Study of Public Policy Implementation on Its Head', in Walter Dean Burnham and Martha Weinberg (eds) *American Politics and Public Policy* (Cambridge: MIT Press), 391–402.

Lipsky, Michael (1980) *Street-Level Bureaucracy* (New York: Russell Sage Foundation).

Lorenz, Chris (1998) 'Can Histories Be True? Narrativism, Positivism, and the "Metaphorical Turn"', *History and Theory*, 37 (3), 309–30.

Mahoney, James and Goertz, Gary (2004) 'The Possibility Principle: Choosing Negative Cases in Comparative Research', *American Political Science Review*, 98 (4), 653–69.

Maynard-Moody, Steven and Musheno, Michael (2003) *Cops, Teachers, Counselors: Stories from the Front Lines of Public Service* (Ann Arbor: University of Michigan Press).

Maynard-Moody, Steven and Stull, Donald (1987) 'The Symbolic Side of Policy Analysis', in Frank Fischer and John Forester (eds) *Confronting Values in Policy Analysis* (Newbury Park: Sage), 248–65.

McDonald, Terrence J. (ed.) (1996) *The Historic Turn in the Human Sciences* (Ann Arbor: University of Michigan).

Miettinen, Reijo, Samra-Fredericks, Dalvir, and Yanow, Dvora (2009) 'Re-Turn to Practice', *Organization Studies*, 30 (12), 1309–27.

Miller, Donald F. (1985) 'Social Policy: An Exercise in Metaphor', *Knowledge*, 7 (2), 191–215.

Minow, Martha (1991) *Making All the Difference* (Ithaca: Cornell University Press).

Müller, Cornelia (2008) *Metaphors Dead and Alive, Sleeping and Waking* (University of Chicago Press).

Nader, Laura (1974/1969) 'Up the Anthropologist – Perspectives Gained From Studying Up', in Dell Hymes (ed.) *Reinventing Anthropology* (New York: Vintage Books), 284-311.

Nicolini, Davide (2009) 'Zooming In and Zooming Out: A Package of Method and Theory to Study Work Practices', in Sierk Ybema, Dvora Yanow, Harry Wels, and Frans Kamsteeg (eds) *Organizational Ethnography* (London: Sage), 120–38.

Oren, Ido (2003) *Our Enemies and US: America's Rivalries and the Making of Political Science* (Ithaca: Cornell University Press).

Pachirat, Timothy (2011) *Every Twelve Seconds: Industrialized Slaughter and the Politics of Sight* (New Haven: Yale University Press).

Pachirat, Timothy (2013/2006) 'We Call It a Grain of Sand: The Interpretive Orientation and a Human Social Science', in Dvora Yanow and Peregrine Schwartz-Shea (eds) *Interpretation and Method*, 2nd edn (Armonk: M.E. Sharpe), 426–32.

Pal, Leslie (1995) 'Competing Paradigms in Policy Discourse', *Policy Sciences*, 18 (2), 185–207.

Polanyi, Michael (1966) *The Tacit Dimension* (New York: Doubleday).

Polkinghorne, Donald (1983) *Methodology for the Human Sciences* (Albany: SUNY Press).

Polkinghorne, Donald (1988) *Narrative Knowing and the Human Sciences* (Albany: SUNY Press).

Pressman, Jeffrey L. and Wildavsky, Aaron (1973) *Implementation* (Berkeley: University of California Press).

Prottas, Jeffrey M. (1979) *People-Processing* (Lexington: D.C. Heath).

Rabinow, Paul and Sullivan, William M. (eds) (1979/1985) *Interpretive Social Science*, 1st and 2nd edn (Berkeley: University of California Press).

Rasmussen, Amy Cabrera (2011) 'Contraception as Health? The Framing of Issue Categories in Contemporary Policy Making', *Administration and Society*, 43 (8), 930–53.

Redman, Eric (2001/1973) *The Dance of Legislation* (Seattle: University of Washington Press).

Rein, Martin (1976) *Social Science and Public Policy* (New York: Penguin).

Rein, Martin and Schon, Donald A. (1977) 'Problem Setting in Policy Research', in Carol Weiss (ed.) *Using Social Research in Policy Making* (Lexington Books), 235–51.

Ricoeur, Paul (1971) 'The Model of the Text', *Social Research*, 38 (4), 529–62.

Roe, Emery (1994) *Narrative Policy Analysis* (Durham: Duke University Press).

Rudolph, Lloyd I. and Rudolph, Susanne Hoeber (2008/1979) 'Authority and Power in Bureaucratic and Patrimonial Administration: A Revisionist Interpretation of Weber on Bureaucracy', *Explaining Indian Democracy: A Fifty Year Perspective*, 1 (2), 168–99.

Rutgers, Mark R. (2004) 'Comparative Public Administration: Navigating Scylla and Charybdis – Global Comparison as a Translation Problem', *Administrative Theory and Praxis*, 26 (2), 150–68.

Sapolsky, Harvey (1972) *The Polaris System Development: Bureaucratic and Programmatic Success in Government* (Cambridge: Harvard University Press).

Schaffer, Frederic Charles (1998) *Democracy in Translation* (Ithaca: Cornell University Press).

Schaffer, Frederic Charles (Forthcoming) 'Thin Descriptions: The Limits of Survey Research on the Meaning of Democracy', *Polity*.

Schatz, Edward (ed.) (2009) *Political Ethnography* (University of Chicago Press).

Schatzki, Theodore R., Knorr-Cetina, Karin, and von Savigny, Eike (eds) (2001) *The Practice Turn in Contemporary Theory* (New York: Routledge).

Schmidt, Ronald, Sr. (2013/2006) 'Value-Critical Policy Analysis', in Dvora Yanow and Peregrine Schwartz-Shea (eds) *Interpretation and Method*, 2nd edn (Armonk: M.E. Sharpe), 322–37.

Schneider, Anne L. and Ingram, Helen (1993) 'Social Construction of Target Populations', *American Political Science Review*, 87 (2), 334–47.

Schneider, Anne Larason and Ingram, Helen (1997) *Policy Design for Democracy* (Lawrence: University Press of Kansas).

Schon, Donald A. (1979) 'Generative Metaphor', in Andrew Ortony (ed.) *Metaphor and Thought* (Cambridge University Press), 254–83.

Schön, Donald A. and Rein, Martin (1994) *Frame Reflection* (New York: Basic Books).

Schram, Sanford F. (2002) *Praxis for the Poor* (New York University Press).

Schwartz-Shea, Peregrine and Yanow, Dvora (2012) *Interpretive Research Design* (New York: Routledge).

Selznick, Philip (1949) *TVA and the Grass Roots* (Berkeley: University of California Press).

Shehata, Samer (2013/2006) 'Ethnography, Identity, and the Production of Knowledge', in Dvora Yanow and Peregrine Schwartz-Shea (eds) *Interpretation and Method*, 2nd edn (Armonk: M.E. Sharpe), 209–27.

Shore, Cris (2011) 'Espionage, Policy and the Art of Government', in Cris Shore, Susan Wright, and Davide Peró (eds) *Policy Worlds* (Oxford: Berghahn), 169–86.

Shore, Cris and Wright, Susan (2011) 'Introduction', in Cris Shore, Susan Wright, and Davide Peró (eds) *Policy Worlds* (Oxford: Berghahn), 1–25.

Shore, Cris, Wright, Susan, and Peró, Davide (eds) (2011) *Policy Worlds* (Oxford: Berghahn).

Soss, Joe (2000) *Unwanted Claims: The Politics of Participation in the U.S. Welfare System* (Ann Arbor: University of Michigan Press).

Soss, Joe (2005) 'Making Clients and Citizens: Welfare Policy as a Source of Status, Belief, and Action', in Anne L. Schneider and Helen M. Ingram (eds) *Deserving and Entitled: Social Constructions and Public Policy* (Albany: SUNY Press), 291–328.

Soss, Joe (2013/2006) 'Talking Our Way to Meaningful Explanations: A Practice-Centered View of Interviewing for Interpretive Research', in Dvora Yanow and Peregrine Schwartz-Shea (eds) *Interpretation and Method*, 2nd edn (Armonk: M.E. Sharpe), 161–82.

Stein, Sandra J. (2004) *The Culture of Education Policy* (New York: Teachers College Press).

Swaffield, Simon (1998) 'Contextual Meanings in Policy Discourse', *Policy Sciences*, 31 (3), 199–224.

Symposium (2009) 'Ethnographic Methods in Political Science', *Qualitative and Multi-Method Research*, 7 (2), 32–50.

Taylor, Charles (1971) 'Interpretation and the Sciences of Man', *Review of Metaphysics*, 25 (1), 3–51.

Torgerson, Douglas (1985) 'Contextual Orientation in Policy Analysis', *Policy Sciences*, 18 (3), 241–61.

Torgerson, Douglas (1986) 'Interpretive Policy Inquiry', *Policy Sciences*, 19 (4), 307–405.

van Hulst, Merlijn (2008) *Town Hall Tales: Culture as Storytelling in Local Government* (Delft: Eburon).

van Hulst, Merlijn and Yanow, Dvora (2014) 'From Policy "Frames" to "Framing"', *American Review of Public Administration* (forthcoming).

Verloo, Mieke (ed.) (2007) *Multiple Meanings of Gender Equality: A Critical Frame Analysis of Gender Policies in Europe* (Budapest: Central European University Press).

Walsh, Katherine Cramer (2004) *Talking About Politics* (University of Chicago Press).

Walsh, Katherine Cramer (2007) *Talking About Race* (University of Chicago Press).

Weatherley, Richard (1979) *Reforming Special Education* (Cambridge: MIT Press).

White, Jay D. (1992) 'Taking Language Seriously: Toward a Narrative Theory of Knowledge for Administrative Research', *American Review of Public Administration*, 22 (2), 75–88.

White, Stephen K. (2004) 'The Very Idea of a Critical Social Science: A Pragmatist Turn', in Fred Rush (ed.) *The Cambridge Companion to Critical Theory* (New York: Cambridge University Press), 310–35.

Yanow, Dvora (1990) 'Tackling the Implementation Problem: Epistemological Issues in Policy Implementation Research', in Dennis J. Palumbo and Donald J. Calista (eds) *Implementation and the Policy Process* (New York: Greenwood), 213–27.

Yanow, Dvora (1993) 'The Communication of Policy Meanings: Implementation as Interpretation and Text', *Policy Sciences*, 26 (1), 41–61.

Yanow, Dvora (1995a) 'Practices of Policy Interpretation', *Policy Sciences*, 28 (2), 111–26.

Yanow, Dvora (1995b) 'Built Space as Story: The Policy Stories That Buildings Tell', *Policy Studies Journal*, 23 (3), 407–22.

Yanow, Dvora (1996) *How Does a Policy Mean? Interpreting Policy and Organizational Actions* (Washington, DC: Georgetown University Press).

Yanow, Dvora (1999) 'Public Policies as Identity Stories: American Race-Ethnic Discourse', in Tineke Abma (ed.) *Telling Tales: On Narrative and Evaluation* (Stamford: JAI), 29–52.

Yanow, Dvora (2000) *Conducting Interpretive Policy Analysis* (Newbury Park: Sage).

Yanow, Dvora (2003a) 'Interpretive Empirical Political Science: What Makes This Not a Subfield of Qualitative Methods', *Qualitative Methods*, 1 (2), 9–13.

Yanow, Dvora (2003b) *Constructing 'Race' and 'Ethnicity' in America: Category-Making in Public Policy and Administration* (Armonk: M.E. Sharpe).

Yanow, Dvora (2007) 'Interpretation in Policy Analysis: On Methods and Practice', *Critical Policy Analysis*, 1 (1), 109–21.

Yanow, Dvora (2012/1992) 'Supermarkets and Culture Clash: The Epistemological Role of Metaphors in Administrative Practice', in Albert J. Mills and Gabriel Durepos (eds) *Case Study Methods in Business Research*, vol. 3 (London: Sage), 123–42.

Yanow, Dvora (2013/2006) 'How Built Spaces Mean', in Dvora Yanow and Peregrine Schwartz-Shea (eds) *Interpretation and Method*, 2nd edn (Armonk: M.E. Sharpe), 368–86.

Yanow, Dvora and Schwartz-Shea, Peregrine (2008) 'Reforming Institutional Review Board Policy: Issues in Implementation and Field Research', *PS: Political Science and Politics*, 40 (1), 483–94.

Yanow, Dvora and Schwartz-Shea, Peregrine (eds) (2013/2006a) *Interpretation and Method: Empirical Research Methods and the Interpretive Turn*, 2nd edn (Armonk: M.E. Sharpe).

Yanow, Dvora and Schwartz-Shea, Peregrine (2013/2006b) 'Introduction', in Dvora Yanow and Peregrine Schwartz-Shea (eds) *Interpretation and Method*, 2nd edn (Armonk: M.E. Sharpe), xiii–xxxi.

Yanow, Dvora and Schwartz-Shea, Peregrine (2013/2006c) 'Doing Social Science in a Humanistic Manner', in Dvora Yanow and Peregrine Schwartz-Shea (eds) *Interpretation and Method*, 2nd edn (Armonk: M.E. Sharpe), 433–48.

Yanow, Dvora and van der Haar, Marleen (2012) 'People Out of Place: Allochthony and Autochthony in Netherlands Identity Discourse–Metaphors and Categories in Action', *Journal of International Relations and Development*, 16 (2), 227–61.

Yanow, Dvora, Schwartz-Shea, Peregrine, and Freitas, Maria José (2009) 'Case Study Research in Political Science', in Albert J. Mills, Gabrielle Durepos, and Elden Wiebe (eds) *The Encyclopedia of Case Study Research*, vol. 1 (Newbury Park: Sage), 108–13.

Zirakzadeh, Cyrus Ernesto (2009) 'When Nationalists Are Not Separatists: Discarding and Recovering Academic Theories While Doing Fieldwork in the Basque Region of Spain', in Edward Schatz (ed.) *Political Ethnography* (University of Chicago Press), 97–118.

Part III

New Challenges in Comparative Designs

8
Comparing Beyond Europe and North America

Joseph Wong

Introduction

The 1982 publication of Chalmers Johnson's book, *MITI and the Japanese Miracle: The Growth of Industrial Policy, 1925 to 1975*, was a watershed moment in comparative policy studies (Johnson, 1982). The book introduced Western readers to the Japanese developmental experience, precisely at the moment when Japan was rivalling the United States in terms of global economic dominance. Johnson's book provided in-depth analyses of trade, investment, human capital development and strategic industrial policies to grow Japanese enterprises into world-class competitors. His study revealed how Japanese firms were organized and how they were linked to the powerful state. The book also explained how policy decisions in Japan were made in a hierarchical manner, how power and authority over resource allocation decisions were centralized in a meritocratic bureaucracy, and specifically within what Johnson called the state's 'pilot agency', the Ministry of International Trade and Industry (MITI). Finally, Johnson's book intimated that a hard or strong state was helpful, at least in Japan's experience, in steering national economic development and continual industrial upgrading.

Though *MITI and the Japanese Miracle* was intended by Chalmers Johnson to be a relatively narrow and focused analysis of Japan's economic development during the middle part of the twentieth century – to be sure, we now know that he was hesitant all along in writing the concluding chapter on comparative lessons to be learned from the Japanese case – the book's impact was and remains extremely far-reaching in terms of comparative analyses of politics and policy (Johnson, 1999). The book's publication was an analytical benchmark because, on one level, it spawned the literature on the Asian developmental state, a purportedly

unique and regionally bounded model of economic policy and style of industrial policy making which extended to other countries in the region. Johnson's book made Asia relevant in comparative discussions about politics and policy. It most certainly made Japan, as a comparative case, more familiar to a new academic and policy audience.

On another level, however, the impact of Johnson's analysis reached far beyond just Japan and even Asia. As it turns out, *MITI and the Japanese Miracle* animated key ideological debates by exploding extant (that is, the Cold War) understandings of state socialism and liberal capitalism, neither of which characterized what Johnson saw in Japan. The book revived comparative discussions about state corporatism in Latin America and Continental Europe (see, for example, Wiarda, 1996); it reminded us of the earlier works of Friedrich List about the nineteenth-century German Zollverein, and Alexander Hamilton's account of early American industrial policies; it forced a serious rethink of World Bank policies at the global level (World Bank, 1993); and it compelled policy decision-makers elsewhere to consider new styles of macro-economic management, modes of strategic trade, public–private coordination and new policies. The book also turned Weber's cultural essentialist notion of the protestant capitalist ethic on its head by invoking Weberian (that is, Western) notions of rational bureaucratic decision-making in a Japanese context (see Wade, 1991; Evans, 1995). Analysis of the Asian miracle launched, on the one hand, critiques of both the dependency and neoliberal conceptions of international political economy, while drawing its own criticisms, on the other hand, from mainstream economists such as Paul Krugman and his characterization of the region's economic dynamos as mere paper tigers (Krugman, 1994). Johnson's book about Japan was, in the end, truly global in scope; it was, implicitly, about *comparing beyond*.

Inspired by the emergence of Asia's relevance in global discussions on politics and policy, and by the importance of Johnson's pioneering work and others like it, this chapter examines the prospects and processes of comparing beyond Europe and North America in the field of policy studies. The following section briefly makes the case in favour of comparing beyond but with an eye towards facilitating a more global dialogue among cases. Comparing beyond, in this respect, is understood to be analytically inclusive globally, rather than separatist in its tendencies in characterizing the Anglo-European world as distinct from cases beyond. *The basic thesis of this chapter asserts that the logic of comparing beyond is not fundamentally dissimilar to the logic of comparing within the Anglo-European world*. While the comparative logic remains similar, it is nonetheless true that the world beyond is unfamiliar for many researchers

of comparative policy studies. The bulk of the chapter, therefore, considers how we might go about realizing a more expansive comparative research agenda, which encompasses both the familiar and unfamiliar. It does so by drawing on existing work featuring Asia and, in turn, highlighting important challenges in truly global comparative policy studies. The illustrative examples centre on qualitatively oriented research though the logic of comparative inquiry and comparing beyond, I suggest, is consistent across both quantitative and qualitative methodologies.

Comparing beyond – Why?

In this day and age, the imperatives for comparing beyond Europe and North America are, I should think, rather self-evident. Having moved past the postcolonial and Cold War moments – and hence the sort of postcolonial curiosity that inspired a comparative political economy of development, or at the time, politically useful notions of first, second and third worlds – we now live in a much more integrated world, culturally, economically and politically. The world is simply a much smaller place. Hence, expanding the scope of comparable cases to include those in the non–Anglo-European world is as much a function of the contemporary processes of modern globalization as it is the result of a supply-side push in the generation of increasingly comparable data, the professionalization (and universalization) of analytical expertise and the transnational appreciation of real world complexities.

However, comparison ought to be purposive. It needs to be structured. It has to be controlled, done smartly. And it has to serve an analytical objective. As Giovanni Sartori once noted, it is not particularly useful to compare rabbits with stones, even if they are, as objects that can be described, technically comparable. It is possible to compare them but, in the end, pointless (Sartori, 1970). At its core, then, the comparative enterprise brings myriad cases and experiences into meaningful dialogue with one another. Comparing beyond is important because it facilitates new analytical encounters among cases that hitherto were understood to be theoretically incomparable or, more practically speaking, difficult to compare. Thankfully, we have moved beyond this analytical parochialism and intellectual provincialism. As Sartori observes, 'comparing is "learning" from the experience of others', adding more forcefully that 'he who knows one country only knows none' (Sartori, 1994: p. 16).

However, the analytical impulse inherent in comparing can, nonetheless, manifest in either greater analytical divisiveness or inclusion, divergent consequences as much the fault of those engaged in comparing beyond as of those who, for whatever reasons, continue to resist the

global trends described earlier. With respect to division, comparing beyond has led many to *a priori* conclude (and even assume) that the non–Anglo-European world is unique unto itself, *sui generis*, and thus necessarily outside a more global debate about politics and policy. According to this logic, places in Asia, Africa and Latin America are *so* different that they, as cases, are theoretically and practically incomparable with the North American and European experiences. A truly global comparative enterprise must, therefore, be intellectually ephemeral, if not impossible – interesting but ultimately useless, like comparing rabbits with stones.

I find these sorts of scholarly worldviews to be problematic, divisive and separatist rather than intellectually inclusive and global. Difference, in my mind, does not mean incomparable; indeed, analytically important differences – such as the Japanese developmental state that was neither capitalist nor state-socialist – can only be surmised by comparing beyond those relatively few cases which have informed the canon and mainstream policy debates.

A more favourable global comparative enterprise in politics, economics, societies and policies encourages the inclusion of new cases and national experiences into the mix that can expand and build upon the existing theoretical canon in comparative politics, or in the specific field of policy studies, innovate beyond the policy status quo. Put another way, the difference between division and inclusion in comparative policy studies is essentially the schism between rejecting and accommodating new experiences within existing policy and theoretical debates. Open-mindedness to the possibility of new concept formation, empirical unfamiliarity, theoretical bias and, in turn, theoretical refinement is a pre-requisite for comparing beyond. Chalmers Johnson's work on Japan, for instance, both spoke to and refined existing theories and concepts of comparative political economy, policy outputs and policy decision-making processes. *MITI and the Japanese Miracle* expanded our collective knowledge about how societies work, in Japan and ultimately elsewhere. Japan, in turn, became a model for emulation, as well as a case for further theoretical and conceptual refinement. The Japanese experience, as recounted by Johnson, was integrated into a more globally inclusive dialogue about politics and policy.

Comparing beyond is useful, if it is done well. Could we imagine the emergence of a 'varieties of capitalism' framework – arguably one of the most influential turns in comparative studies of modern capitalism and economic policy in recent decades – without the insights gained from comparing Japan and Germany with the United States and the United

Kingdom? (Hall and Soskice, 2003). Can we understand the technological rise of Ireland or Israel without reference to the postwar Asian experience? (Breznitz, 2007). Can we explain judicial activism in public policy, what Ran Hirschl refers to as juristocracy, in young democracies such as South Africa in the absence of comparative insights from the American and Canadian experiences in constitutional reform? (Hirschl, 2004). Does it surprise us that scholars of Canadian federalism provide policy advice to places such as South Africa, India and Afghanistan? Could we fully appreciate the unique governance mechanisms and policy-making institutions of the European Union without reference to other forms of regionalism in Asia, Africa or Latin America? Can we imagine, moving forward, multilateral policy-making processes at the global level without some sense of how decisions are made in China, India, Brazil and Russia? I think, unlikely.

Comparing beyond – How?

Comparing beyond Europe and North America is really not that complicated. Sartori (1994: p. 16), echoing many others who favour more inclusive comparative work, explains concisely that 'comparing is controlling', and that comparative analysis essentially involves 'comparative checking'. If we understand the comparative enterprise in this way, then, comparing *within* Europe and North America and comparing *beyond* effectively employ the same logic of causal inference, the same rules of comparative research design, the same standards for conceptual clarity, the same methods of empirical observation and so on. Whether comparing within or beyond, we define core concepts (such as democracy, bureaucracy, public policy and so on), operationalize them, control for differences and similarities among a spectrum of cases and, in so doing, deductively test sensible hypotheses which explain outcomes. And when such comparative checking fails to confirm what we expect, we inductively generate plausible alternative explanations and formulate new conceptual categories such as Johnson's developmental state. The core logic inherent in comparing education policy processes in the United States and Canada, therefore, is fundamentally no different than the logic of comparing family policy decision-making processes in Japan, Korea, Canada and Italy, or industrial technology policy in Taiwan, Israel and Ireland.

Dealing with unfamiliarity

What is different, however, is that doing comparisons among cases within Europe and North America is *familiar*; these cases, for historical

and geopolitical reasons, make up the theoretical canon and the public policy status quo. They are standard reference points, whether, for example, in terms of new public management (the United States) or parliamentary policymaking (the United Kingdom) or the welfare state (Sweden) or social insurance policies (Germany) or corporatist industrial policy arrangements (Austria) and so on. This is not to say that the universe of comparable cases within Europe and North America are homogenous; obviously they are not. But the differences among them are nonetheless familiar enough to those working in the Anglo-European or Western academy so as to make these cases more easily comparable. In short, their apparent comparability is because of their familiarity.

There is, without doubt, a historical familiarity with key Western cases such as the United Kingdom, the United States, Germany, Canada, France and so on. Histories, for instance, are inextricably tied. Policy diffusion and common democratic politics engender familiarity. Spatial-geographical contiguity as well as ideological convergence and a shared political heritage help blunt or make sense of national differences. Moreover, as key cases in the theoretical canon, mainstream social science theory and the concepts that are the foundations of this theory, allow us to control, explain and ultimately accommodate in our comparative work important variations among and within the Anglo-European experiences. For instance, Putnam's (1994) explanation for regional differences in civic engagement in northern and southern Italy hinges on the familiar concept of social capital; Esping-Andersen's (1991) 'three worlds' of welfare capitalism are ultimately shaped and determined by generally agreed-upon conceptions of social class, social mobilization and democratic-electoral politics just as theories of congressional decision-making in the United States are shaped by the well-tread concepts of principals and agents.

In other words, though the logic of comparing cases within and beyond Europe and North America is consistent, I contend, it is also true that the historical, geographical and theoretical closeness of Western cases seems to make comparisons within (rather than beyond) more familiar and thus more plausible than comparing outside what is familiar. Comparing beyond may not be as meaningless as comparing rabbits with stones, as Sartori warns, but for many, comparing beyond might nonetheless seem like comparing apples with oranges, when comparing apples with apples is more familiar, more intuitive. Therefore, what makes comparing beyond seemingly challenging, is not any inherent incomparability of nations and cases beyond the Anglo-European world, but rather the unfamiliarity we experience when it comes to comparative

policy studies of Asia, Africa and Latin America or for that matter any cases that we know less about.

To reiterate: the obstacle to comparing beyond is unfamiliarity, not incomparability. To make my point more clearly, let me draw again on the notion of comparing apples and oranges. The usage of the axiom 'it's like comparing apples and oranges' seems to suggest that apples and oranges cannot be compared, or in the least, such a comparison would not be particularly useful. I disagree. Their comparative encounter, if we can think of it in those terms, might be one of unfamiliarity and difference, but certainly we can still compare apples and oranges and, in so doing, draw important conclusions or inferences (important, obviously, if one is to choose one to eat). Shape, size, colour, taste, type of fruit, and so on, are all potentially meaningful bases upon which to compare (depending on what one is interested in concluding). My point is that comparing within and beyond – be it different kinds of fruit or different countries – is both possible and revealing provided we first generate and then employ clear, meaningful and consistent concepts and measures, and avoid what Sartori (1970) calls 'conceptual stretching'. To do so, we need to know what it is we are comparing.

As a practical exercise, comparing beyond, to make the unfamiliar more familiar, involves many practical obstacles or hurdles that need to be cleared. Testing hypotheses and generating alternative explanations in non–Anglo-European settings normally require fieldwork – significant amounts of time in the specific field site outside the West. Researchers thus need to become immersed in new cultures, societies and political and economic systems. To engage in field work in places such as Asia, Africa and Latin America also requires field researchers to acquire, for instance, language skills, which are typically 'less commonly taught languages' (a term used at the University of Toronto) or a part of otherwise highly specialized programmes in few universities and colleges in Europe and North America. Researchers need to learn about new sociocultural settings, which can be vastly different from those in the West, in order to effectively carry out fieldwork. Scholarly and professional networks have to be established. Sources of funding need to be discovered. Put another way, gaining the needed skills and resources to compare beyond and to mitigate the unfamiliarity of cases beyond the Anglo-European world is a significant investment.

The practical task of collecting data is problematic too. Integrating new cases into a more global comparative enterprise requires the generation of new data, broadly defined, by which I mean both quantitative and statistical data as well as more qualitative forms of empirical

evidence. Until Johnson's (1982) publication of *MITI and the Japanese Miracle*, little was actually known about how Japanese decision-makers made economic policy because of the dearth of empirical data about that particular case. Comparing beyond means making up a data deficit, which can be, practically speaking, a daunting task.

And the generation of new data alone is not enough. The field of comparative policy studies also requires that data be reliable, that it be replicable and unbiased. Survey research design, for example, is very difficult in non-Western settings, in part due to the dearth of professional polling agencies but also because of the lack of experience potential respondents have with standardized formal surveys and the sort of (impersonal and iterated) interactions that accompany such forms of data collection. Gaining access to key decision-makers for interviews is also difficult in settings where such forms of elite interaction are not encouraged or in non-democratic contexts where efforts at political transparency are highly suspicious. Finally, collecting data that is easily comparable is challenging, particularly when concepts, measurements and indicators are not standardized across cases. My research on biotechnology development in Asia demonstrates, for example, just how elusive it is to generate the most elementary set of comparable indicators of R&D spending in the field of biotech, never mind a consistent definition of what even constitutes 'biotech' in the first place in otherwise relatively similar Asian cases (Wong, 2011).

When confronted with the research challenges of comparing beyond and comparing with the unfamiliar, practical considerations are very important. As a researcher, one must have a much bigger scope of analysis, a wide-angle lens through which to see the empirical world. Details are thus necessarily lost. Missing data is a perennial problem. Case-study work is uneven in terms of empirical depth. Bigger inferential leaps are also required. As a researcher, one must be aware of these challenges, and where possible, compensate or mitigate their effects. In some of my own work, for instance, I have had to present more aggregated quantitative data, when in fact more finely tuned but comparable data would have demonstrated more forcefully my argument. I have had to use translators when interviewing informants in non-English and non-Chinese speaking societies (such as in Korea, India and parts of China). I have had to rely on more 'macro' explanations because my research findings have occasionally lacked the micro-level details to offer such an explanation. I have had to make comparative inferences in cases about which I am less familiar. These are among the many trade-offs any researcher has to make when comparing beyond. Practical issues of comparing

beyond are formidable. Yet, with time and the accumulation of research experience in cases beyond Europe and North America, the practical challenges I have described here can gradually be overcome.

But there are nonetheless deeper concerns, which the rest of this chapter focuses on, that ought to be flagged if we look to develop a more inclusive, global comparative research enterprise in the field of policy studies. These concerns reflect ontological assumptions about how societies beyond the European and North American mainstream work in reality, and how such fundamental differences and unfamiliarity with such cases impact comparative research. *What we might expect to see in countries and cases beyond might not, in the end, jibe with what the researcher actually observes*. This ontological gap between expectations and reality – which I suggest is most evident in conceptions of rationality, heuristic biases and the application of mainstream social scientific theory – should not preclude or discourage comparing beyond; rather, this ontological gap, I argue in the ensuing sections, can in fact reveal important insights in comparative politics and policy studies, provided the researcher is willing to see them. Comparing beyond requires that we relax some of our extant ontological assumptions about how different places work in reality and, in turn, accommodate such experiences into our existing social science theories, concepts and comparative analyses of public policy.

Contextualized rationality

Policy decision-makers are assumed to behave rationally. End users or the targets of public policy instruments and incentives are expected to react rationally as well. Policy decisions are thus generally understood to be a function of rational decision-making processes involving the acquisition and interpretation of information and the subsequent cost-benefit considerations of such information according to relatively consistent and known preferences. I have no objection to this broad conceptualization of rationality nor to the notion that under most circumstances, people behave rationally. However, rationality and rational decision-making need to be contextualized, particularly when considered in unfamiliar cases beyond the European and North American settings.

Peasant life in poor and rural Southeast Asia, the focus of Scott's (1977) study, *The Moral Economy of the Peasant: Rebellion and Subsistence in Southeast Asia*, is an unfamiliar setting. And while Scott's study does not contradict the notion of rationality, his book does highlight the analytical paucity that comes with applying non-contextualized or universalist assumptions of market-based rationality to explain peasant decisions in

precolonial Burma and Vietnam (Scott, 1977). In the book, Scott reveals what he calls a 'subsistence ethic' shared among Southeast Asian peasants, a value system which mirrors those found in precapitalist agrarian societies in Europe. He argues that the chronic 'fear of food shortages' means that, for peasants, their priority is mere survival amidst precarious and seasonal agricultural output. The pursuit of surplus accumulation and individual profit, therefore, is not an optimal strategy for poor peasants who are 'living so close to the margin'. Scott explains that the 'logic of the subsistence ethic' shapes the 'choices and values of much of the peasantry in Southeast Asia' (Scott, 1977: p. 2, 55). Thus, contrary to market expectations of rational behaviour and individualistic rational decision-making, peasants in precolonial Southeast Asia behaved according to deeply entrenched norms of community reciprocity and legitimacy, and pursued strategies for dealing with short- and long-term risk, which together underpin what Scott sees to be the basis of a moral economy.

Scott's observations were not without critics, however. Popkin's (1979) study, *The Rational Peasant: The Political Economy of Rural Society in Vietnam*, critically addresses Scott's findings. Applying the rational choice model of decision-making, Popkin, in contrast to Scott, suggests that peasants in Vietnam behaved in individually self-interested ways, the implication being that Scott's moral economy, or what he observed, must be a function of irrational behaviour on the part of peasants. One interpretation of Scott's work, however, is that peasants are not irrational, even if for Popkin their behaviour cannot be explained in any other way. For Scott, peasant decisions aimed at mitigating the near-term risks of rural life rather than the accumulation of long-term gains potentially accrued from surplus agricultural production are in fact perfectly rational when considered against the realities experienced in rural peasant settings. Rationality, Scott teaches us, is therefore context-specific.

Social science theory does, in some ways, accommodate a more flexible application of rationality. Indeed, the notion of 'bounded rationality' suggests that otherwise rational decision-makers routinely face problems of incomplete information, inconsistent preference orders as well as individuals' cognitive limits in processing complex information (Jones, 2001). It seems to me, however, that what Scott is describing in *The Moral Economy of the Peasant* is not bounded rationality, as it is understood conventionally, but instead a decision calculus – or as Scott puts it himself, a 'logic' – that is constituted by the existential realities of poor peasant life in rural Southeast Asia. The logic of Scott's moral economy thus reflects fundamentally different ways in which societies might

work outside the Anglo-European world, especially societies that are very poor and which might hold to different cultural values. These differences need to be accounted for when comparing beyond.

Bringing societies in Asia, Africa and Latin America into the broader more global comparative enterprise, for instance, requires analytical attention to how rational decision making is shaped by the existential realities of the *very* poor. Economic experiments tell us, for example, that farmers in India who rely on seasonal wages are far more inclined to save if their earnings are paid out in numerous instalments rather than lump sums, regardless of any interest-based inducements which are typically used to encourage saving elsewhere (Soman, 2010). Field research in China reveals that workplace public health education campaigns are less effective, not because workers are less inclined to participate in such campaigns but rather because of labour market volatilities and thus shorter work tenures in any one place (Qian et al., 2007). Cultural differences matter as well in informing rational behaviour. We know that male-dominant gender relations in traditional societies tend to determine women's health decisions even more than medical need or the cost of care (Shaikh and Hatcher, 2004). We also know that the key determinants of health-care utilization in poor countries include not only cost and availability but also societal 'acceptability', or the extent to which health-care interventions meet 'the social and cultural expectations of individual users and communities' (Peters et al., 2008). In other words, decision-makers in these non–Anglo-European societies behave rationally, but, as these examples suggest, what constitutes rational in Asia (and presumably Africa and Latin America) is conditioned by different socioeconomic and cultural contexts.

Thus, 'good value propositions' from the perspective of the rational decision-maker, need to be treated as variables. For example, the conventional wisdom about which factors affect health-seeking behaviour rests almost exclusively on the supply of medical care (that is, availability of services) and cost (that is, financial capacity of the end user). Demand-side considerations are essentially a given. After all, it is inconceivable that anyone, provided that care is readily available, would reject health care services; to do so would be irrational behaviour as a purely deductive understanding of the problem would tell us. Comparing beyond – or more generally comparing with what is unfamiliar – requires us to be open to more inductive, and thus potentially alternative, explanations.

My current research in rural India suggests, for example, that health-care decisions are shaped by the existential realities of poor, rural village life, and that rational decisions are context-specific. Take, for example,

the use of diagnostic screens in rural health-care clinics. Evidence-based research tells us that the early adoption of diagnostic screens is a cheap and cost-effective way to diagnose and treat potentially expensive long-term illnesses. Simply put, early screening not only makes good economic sense to the health-care system, but it also makes good medical sense to the individual. And yet, utilization rates of diagnostic screens among village dwellers in southern India remain relatively low. My research reveals that, for very poor villagers, the cost (as minimal as it is) of screening is still onerous, particularly given that the outcomes of diagnostic screens are uncertain (which is why they are unwilling to spend any money on a test to determine that there is nothing wrong with them). Moreover, the additional transaction costs of health-care utilization – such as the cost of transportation, lost time from work and so on – are burdensome for poor villagers, even if the eventual costs for the patient would be unequivocally catastrophic if an illness or condition were to go undetected over a longer period of time. In other words, poor villagers in rural Southern India view inexpensive diagnostic screens to be a relatively poor value proposition given the perceived socio-economic risks, resources and volatilities of peasant life. From the perspective of the outside researcher, such decisions might appear irrational, though when put into context, they make undeniable sense from the perspective of the poor peasant.

Rationality, understood broadly, is universal in that all decision-makers behave rationally, including, as I have argued, Scott's precolonial peasants in Southeast Asia or Indian villagers today. But rationality in practice is determined by particular socio-economic and cultural contexts. Being very poor, for example, entails a different kind of rationality and involves a profoundly different kind of decision calculus. In terms of policy, then, recognizing the context-specific bases of rational decision-making ultimately leads to better policy design in settings and cases that are unfamiliar beyond Europe and North America. On a more conceptual and theoretical level, recognizing rationality to be shaped by context-specific realities permits the ontological accommodation that is necessary if we are to meaningfully compare beyond.

Heuristic cues

Policy making is about choosing from among a range of policy instruments. Political battles and bureaucratic wrangling in the policy process *everywhere* reflect the contestation of 'causal stories' that defines contending policy alternatives (Stone, 1989). Kingdon's study (1984) on the politics of policy agenda setting similarly describes how policy entrepreneurs clash over policy problems and the appropriate solutions

to those problems. We also know that policy advocates use heuristic cues to help make sense of complicated and uncertain policy debates. In other words, heuristic cues – such as ideologies, best past practices, efficiency goals – simplify the world. They funnel choices in otherwise complex decision-making processes. For instance, we are familiar with the asserted notions that markets lead to systemic efficiencies; publicly financed social insurance is more effective at re-distribution; increased household savings tend to increase domestic investment; private for-profit health insurance carriers tend to cream-skim among potential enrolees; intellectual property rights protection facilitates innovation; the achievement of universal education corresponds with industrial upgrading; and so on. These cues, which are subjectively generated and biased, nonetheless allow us to navigate complicated policy decisions.

By helping us to navigate policy alternatives, these heuristic cues – or biases – shape policy debates and policy processes as well as inform policy decisions. Awareness of and familiarity with these cues and biases, from the perspective of the comparative policy studies researcher, provides an important analytical lens through which to understand how policy decisions are made both here and beyond. Indeed, *heuristic cues and biases are context-specific*, generated from certain locational, temporal and experiential contexts. Put another way, heuristic cues and biases 'are normative frames derived from national experiences and repertoires of proven decisions internalized from the past'. (Wong, 2011: p. 177).

For example, my research suggests the heuristic cues and biases entrenched in the Asian developmental state model have continued to shape current economic policy debates in the region even though policy decision-makers claim to have moved beyond the postwar developmental strategy of picking and making industrial winners. The story I tell in *Betting on Biotech: Innovation and the Limits of Asia's Developmental State* explains how deeply embedded heuristic biases, generated from the experiences of the region's past, have persisted into the current era of the knowledge economy despite both wilful efforts of decision-makers and the functional imperatives of science-based industrialization to evolve away from an earlier top-down style of policy decision-making of the developmental state model. I argue in the book that to make sense of and explain policy processes and decisions in East Asia requires prior knowledge of the region's past and its heuristic imprint on contemporary debates. To do so, I had to comb through archives, government documents and industry reports and conduct hundreds of stakeholder–informant interviews in Taiwan, Korea, Singapore, China, Hong Kong and the United States.

Beginning in the 1990s, the governments in Taiwan, South Korea and Singapore announced their intentions to steer their already advanced economies towards more innovation-driven industrial activities. Biotechnology was flagged in all three cases to be a key sector in the new economy. Since that time, billions of dollars have been invested in the sector. Meanwhile, new R&D institutions have been built, universities reoriented and new industrial policies put into place. The basic argument in *Betting on Biotech* is that the emergence of the new innovation-driven economy in Taiwan, Korea and Singapore has consequently resulted in the retreat of the developmental state. Increasingly incapable of handling the uncertainties of science-based industrialization, regulation and commercialization, the state has, in many significant ways, given way to the initiatives of the private sector and industry specifically. Private sector R&D spending has increased manifold, now accounting for the majority of research funding in all three places. R&D agendas and consortia are increasingly organized by firms, rather than from the top-down, which had been the case during the 1990s. Technology transfer is handled by firms and labs rather than brokered by the state. Similarly, regulatory leadership has emerged from outside the state apparatus, involving academic researchers, regulatory experts and industry. In other words, we have seen over the past 15 years or so *the retreat of the developmental state* and specifically its role in targeting and making, from the top-down, industrial winners. What is more, decision-makers in Taiwan, Korea and Singapore have understood this industrial policy turn to be pointed in the right direction moving forward.

And yet, despite what I observe in the book to be a general trend away from the strategic logic of the developmental state, industrial policy decision-makers in Taiwan, Korea and Singapore have continued to pursue a strategy of manufacturing or creating 'stars'. Though not as conspicuous as in the past, the governments in Taiwan, Korea and Singapore have, on a much smaller though still significant scale, continued to over-invest in key targets with the explicit aim of creating a biotechnology star. Government investment in certain Taiwanese life sciences firms, over-investment in star scientists such as Korea's stem cell researcher Hwang Woo-Suk, protecting and subsidizing key firms such as we see in Singapore – these are all examples of where state-led strategies for industrial development of the past have persisted into the present, despite a purported shift away from the developmental state model. How can we explain this?

Continuity, in these respects, is a function of path dependency. On one level, the strategy of manufacturing biotech stars reflects an institutional path dependency whereby key decision-makers and institutional

constraints have been held-over from an earlier period. As we know, institutional change, regardless of actors' intentions or efforts to effect change, is a slow process. But on an even deeper level, respondents in Taiwan, Korea and Singapore recount that, despite an overall strategic re-orientation of the state and its role in promoting industrial develop-ment, the legacies or heuristic imprints of the post war developmental state era have persisted and have in turn shaped how decision-makers in Asia approach the challenges of industrial upgrading. Past practices and proven strategic repertoires in industrial development, as explained to me, have been hard to shake. Heuristic biases in favour of heavy-handed state leadership continue to *make strategic sense to decision-makers* even as they are aware of the limits of such an approach.

My point is that in order to understand how policy decisions are made and why certain policy instruments are chosen over others, we need to identify the heuristic cues and biases which inform policy debates and contestation, and in so doing help determine policy choices. That decision-makers in Taiwan, Korea and Singapore have chosen to manufacture stars in the biotech sector is not my main point. Rather, it is that to understand why and how they have chosen such a strategy at all, despite important trends away from such choices, one must appreciate the lega-cies of the developmental state, as both institutional constraints and as ideational path dependency, wherein certain 'values, beliefs, cognitive cues and mental scripts' historically generated within these particular societies continue to inform how decisions are made (Wong, 2011: p. 177). Analytically, therefore, comparing beyond Europe and North America requires the researcher to account for heuristic bias and differ-ences in order to explain how decision-makers in settings beyond both understand and resolve policy problems and their solutions. In short, analyzing policy processes in non–Anglo-European settings needs deep knowledge of unfamiliar pasts and also some sense of how they shape the present.

Theoretical bias

Just as heuristic cues help decision-makers clarify complex policy problems and policy solutions, social science theory aids the researcher in simpli-fying explanations about otherwise complex policy decision-making processes. Theory is the basis of the comparative enterprise. Analysts essentially 'hang' their empirical evidence on theoretical frameworks to generate explanatory power for their findings. Theory is powerful because it points our analysis in certain directions, and it shapes the conclusions we draw. But theory, similar to those heuristic cues used by

decision-makers to help navigate policy conflict described above, is biased. Social science theories are derived from national experiences and particular contexts. As a result, *how* we employ theory when comparing beyond Europe and North America can either obfuscate or illuminate our empirical findings. Bias in itself is not the problem. Rather, awareness of theoretical bias is critical. Gaining familiarity with cases beyond ensures that theoretical bias is accounted for.

The theory of the modern welfare state has proven to be extremely resilient in the social science mainstream. As a theoretical framework for comparatively understanding social policies, the theory of the modern welfare state has also 'travelled' remarkably well from case to case. The pioneering work of Walter Korpi and Gøsta Esping-Andersen, which argued that the formation of welfare states was caused by working class mobilization, has continued to inform theoretical debates about social policy reform in the developed and developing worlds (Korpi, 1983; Esping-Andersen, 1991; Haggard and Kaufman, 2008). Comparative analysis of the welfare state has developed measures of the redistributive effect of social policy and indicators of welfare state commitment; key assumptions about the political and economic requisites of welfare state formation; and elegant theoretical assertions regarding the political mechanisms, such as the power resources available to the organized working class, at work in shaping social policy regimes. Indeed, power resources theory and the politics of working class mobilization have gone a long way in explaining, why, for instance, we see well-developed welfare states in parts of continental Europe and why the welfare state has failed to take hold in other countries. The works of Korpi, Esping-Andersen, Huber and Stephens, Pierson and many others have also shed light on why we see variations in the political economy of welfare state formation, evolution and retrenchment, and thus different kinds of social policy regimes around the world (see, for example, Pierson, 1995; Huber and Stephens, 2001).

As much as power resources theories of the welfare state have been generally accepted as the conventional wisdom in explaining social policy-making processes, they have also been misleading with respect to some of the conclusions that have been drawn about social policy processes in other, more unfamiliar cases beyond the European and North American canon. Gregory Kasza, for instance, in his powerful study, *One World of Welfare: Japan in Comparative Perspective*, argues that Esping-Andersen's 'three worlds' framework – and its emphasis on social spending, the political strength of the left and the timing of social policy reform – leads one to incorrectly characterize Japan as a welfare state laggard, as

is often the case in the welfare state literature (Kasza, 2006). Kasza shows in his book that, in fact, Japan's social policy mix is very much in-line with Western OECD countries, despite comparatively low levels of social spending and a relatively weak left in party politics. He also demonstrates how the political evolution of social policy in Japan adhered to similar patterns of development in other archetypal Western welfare states, thus de-bunking the idea of Japanese 'welfare society'. Kasza's work, among the most informative of Japanese policy studies, shows how theoretical bias can mislead, and that, if one looks to meaningfully compare beyond, researchers must be aware of this bias and refine their theoretical toolkits to accommodate new cases outside of the European and North American mainstream.

The analysis of social policy development in other parts of industrial East Asia, such as Taiwan and South Korea, reflects this theoretical bias and thus the imperatives of gaining familiarity of cases beyond if we intend to learn from globally comparative policy studies. According to the power resources theory of the welfare state, the strength of the political left – measured by labour union density and the electoral popularity of left-leaning or social democratic political parties – is directly related to the re-distributive capacity of the welfare state and public social spending. The power of the left helps us predict the emergence (or not) of the welfare state. In places such as Taiwan and Korea, however, labour unions are politically very weak, unionization rates are low and working class identity is weak and fragmented. Furthermore, there are no social democratic parties; the legacies of the Cold War have made it difficult for the emergence of ideologically left-leaning parties in most of Asia. In short, there is no identifiable 'left' in industrial East Asia, a fact that has led many to conclude (incorrectly, in my view) that there cannot be a welfare state in places such as Taiwan and Korea. Some have argued that social policy in the Asia region is 'productivist', intended only to support economic growth rather than equitable re-distribution (Holliday, 2000). Others have asserted that social protection and social safety nets are provided primarily through family and kinship networks, reflective of deep cultural norms supposedly unique to Asian societies (Jones, 1990). In the absence of a political left, these explanations are made even more compelling. A purely deductive reasoning would bias us to conclude that the welfare state does not – because it 'theoretically' cannot – exist in East Asia. Inductive reflection, as a correction to such theoretical bias, potentially tells us something else.

My take on the region, for instance, contends that the inherent biases of conventional welfare state theories obscure the realities of social

policy development in East Asia. An overly rigid reading of power resources theory points us in the wrong direction, contributing to what I see to be misleading conclusions about the evolution of social policy regimes in the region. It fails to highlight what is actually going on in the region. Peng and Wong (2008) argue that publicly managed social insurance programmes in health, old-age income security and social care in Taiwan and South Korea are re-distributive and that, furthermore, their political evolution has been motivated by increasing their re-distributive capacity. We argue that the 'purpose' of social insurance and continued institutional reform is to enhance socio-economic equity and re-distribution; their commitment to the welfare state is deepening.

In my book, *Healthy Democracies: Welfare Politics in Taiwan and South Korea*, I develop an alternative political explanation for the decisions to universalize health insurance in Taiwan and Korea during the 1990s, thus refining the theoretical conventional wisdom found in power resources theory (Wong, 2004a, b). Recognizing the absence of a strong political left, particularly in Taiwan's and Korea's political party system, I argue instead that it was the formation of cross-class coalitions that compelled policy decision-makers there to pursue equity-enhancing reform. The evidence also suggests that progressive social policy advocates were primarily positioned within civil society, rather than the party system, which gave legislators the ideological flexibility to form political–electoral coalitions without being labelled as leftist parties and politicians. Unlike in Latin America, what we saw in Taiwan and South Korea were progressive social forces taking advantage of a weakly institutionalized left in order to push, from the bottom-up, re-distributive social policy reforms (see also Wong, 2004a, b). The welfare state was thus created because of, rather than despite, the absence of a strong political left. Theory would predict that there ought to be a stronger political impetus for welfare state expansion in Latin America than in East Asia when, in reality, this is not the case.

My work on East Asian welfare states does not dismiss the theoretical insights of power resources theory nor does it reject the possibility of welfare state formation in Taiwan and South Korea. Rather, in comparing beyond, my inductive analysis of social policy decision-making and social policies in East Asia accounts for theoretical bias and, in so doing, refines rather than rejects conventional social science theory. In short, the work I have briefly outlined here shows not only that it is possible to integrate cases beyond Europe and North America, but also how such analyses can build upon and, in turn, further social science theory aimed at explaining policy processes and outcomes when comparing beyond.

Conclusion

To compare beyond requires the analyst to account for different kinds of rationality, heuristic cues shared among decision-makers and theoretical biases, which can lead to incorrect conclusions. In the main body of this chapter, I have demonstrated how rationality is context-specific and that seemingly irrational decisions can in fact be quite sensible given different and unfamiliar circumstances. I have also demonstrated how, in settings beyond Europe and North America, policy decision-makers and policy processes are shaped by enduring heuristic cues that reflect different histories, values and cognitive scripts; in short, an appreciation of antecedent decisions, pathways and strategic choices is critical for explaining subsequent decisions. And finally, I have shown how the application of social science theory – though with an awareness of theoretical bias – particularly those theories which endeavour to explain policy decisions and policy processes, can correct misleading conclusions and, in turn, reveal theoretically plausible alternative explanations.

In summary, there are four key points one must consider when comparing beyond Europe and North America. First, the researcher needs to take care of the many *practical obstacles* which might emerge when comparing cases that are unfamiliar and beyond, including field research in order to understand (at least minimally) particular contexts; language training, especially if he or she intends to interview native-language speaking informants and consult primary materials in the field; the creation of a network of contacts in the various cases beyond; and the resources needed to conduct the research. Second, one must approach the comparative study with *conceptual clarity* (that is, what constitutes a democracy, what is a welfare state, how to measure a successful industrial policy, and so on) and thus a consistent basis of comparison. Conceptual clarity is imperative for both qualitative and quantitative research methodologies. Third, and related to my earlier point, the researcher needs to employ a deductive logic of inference to test, either implicitly or explicitly, prevailing hypotheses and explanations about certain outcomes (this is usually done, minimally, in the research design stage). But in order to avoid what I have described in this chapter as theoretical bias, he or she should also be prepared to *inductively generate new concepts and refine existing theories* to reflect the empirics of the cases being considered. Put another way, comparing beyond and with the unfamiliar must avoid what Sartori famously derided as 'conceptual stretching', the ladder of conceptual abstraction that ultimately runs the risk of making conceptual categories into meaningless bases of

comparative enquiry. Finally, and this is more of an ontological caution for the researcher looking to compare beyond: *respect unfamiliarity* by appreciating real-world complexities beyond the familiar and the (likely) fact that places beyond might not work in the ways that he or she expects them to.

The four summary points I have outlined earlier are precisely those which we would likely consider when doing any comparison. In this way, then, the principal argument of this chapter falls very much in-line with mainstream ways of doing comparative research on *all* societies, politics and public policies. To be sure, I caution against ontological extremes: that analysts, on the one hand, need to avoid being too universalistic or homogenizing when comparing different societies (that is, that rationality is narrowly and universally applicable) and, on the other hand, must also move away from the sort of 'other-world' essentialism that characterizes unfamiliar cases beyond Europe and North America as *a priori* incomparable. Analysts should not assume that non-Western cultures, societies, polities and thus their policies and policy processes are so fundamentally foreign that comparing beyond is ephemeral – like comparing rabbits with stones. Both of those perspectives are problematic. The way to cut through either of these two extreme ontological positions, I have argued, is to recognize the inherent challenges of *overcoming unfamiliarity*, of comparing unfamiliar cases and national experiences. Deductive reasoning illuminates the limits of general theory, which in turn invites more inductive forms of analyses for refining or generating alternative explanations. Put another way, we do not need to 're-invent the wheel', because, as I have argued here, conventional social science approaches to comparative analyses already allow like-minded scholars to become familiarized with different cases, to recognize how differences matter in terms of concepts and theories and to bring different experiences into dialogue with one another. Comparing beyond, therefore, is to become familiar with what is beyond.

Bibliography

Breznitz, Dan (2007) *Innovation and the State: Political Choice and Strategies for Growth in Israel, Taiwan and Ireland* (New Haven: Yale University Press).

Esping-Andersen, Gøsta (1991) *The Three Worlds of Welfare Capitalism* (Princeton University Press).

Evans, Peter (1995) *Embedded Autonomy: States and Industrial Transformation* (Princeton University Press).

Haggard, Stephan and Robert Kaufman (2008) *Development, Democracy and Welfare States: Latin America, East Asia and Eastern Europe* (Princeton University Press).

Hall, Peter and David Soskice (eds) (2003) *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage* (New York: Oxford University Press).

Hirschl, Ran (2004) *Towards Juristocracy: The Origins and Consequences of the New Constitutionalism* (Cambridge: Harvard University Press).

Holliday, Ian (2000) 'Productivist Welfare Capitalism: Social Policy in East Asia', *Political Studies*, 48 (4), 706–23.

Huber, Evelyne and John Stephens (2001) *Development and the Crisis of the Welfare State* (University of Chicago Press).

Johnson, Chalmers (1982) *MITI and the Japanese Miracle: The Growth of Industrial Policy, 1925–1975* (Stanford University Press).

Johnson, Chalmers (1999) 'The Developmental State: Odyssey of a Concept', in Meredith Woo-Cumings (ed.) *The Developmental State* (Ithaca: Cornell University Press).

Jones, Bryan (2001) *Politics and the Architecture of Choice: Bounded Rationality and Governance* (University of Chicago Press).

Jones, Catherine (1990) 'Hong Kong, Singapore, South Korea and Taiwan: Oikonomic Welfare States', *Government and Opposition*, 25 (4), 446–62.

Kasza, Gregory (2006) *One World of Welfare: Japan in Comparative Perspective* (Ithaca: Cornell University Press).

Kingdon, John (1984) *Agendas, Alternatives and Public Policies* (New York: Longman).

Korpi, Walter (1983) *The Democratic Class Struggle* (Abingdon: Routledge)

Krugman, Paul (1994) 'The Myth of Asia's Miracle', *Foreign Affairs*, 73 (6), 62–78.

Peng, Ito and Joseph Wong (2008) 'Institutions and Institutional Purpose: Continuity and Change in East Asian Social Policy', *Politics and Society*, 36 (1), 61–88.

Peters, David et al. (2008) 'Poverty and Access to Health Care in Developing Countries', *Annals of the New York Academy of Science*, 1136 (1), 161–70.

Pierson, Paul (1995) *Dismantling the Welfare State?* (New York: Cambridge University Press).

Popkin, Samuel (1979) *The Rational Peasant: The Political Economy of Rural Society in Vietnam* (Berkeley: University of California Press).

Putnam, Robert (1994) *Making Democracy Work: Civic Traditions in Modern Italy* (Princeton University Press).

Qian, Xu et al. (2007) 'Promoting Contraceptive Use Among Unmarried Female Migrants in One Factory in Shanghai: A Pilot Workplace Intervention', *BMC Health Services Research*, 7 (77), 17–22.

Sartori, Giovanni (1970) 'Concept Misformation in Comparative Politics', *American Political Science Review*, 64 (4), 1033–53.

Sartori, Giovanni (1994) 'Compare Why and How: Comparing, Miscomparing and the Comparative Method', in Mattei Dogan and Ali Kazancigil (eds) *Comparing Nations* (Oxford: Blackwell).

Scott, James (1977) *The Moral Economy of the Peasant: Rebellion and Subsistence in Southeast Asia* (New Haven: Yale University Press).

Shaikh, Babar and Juanita Hatcher (2004) 'Health Seeking Behaviour and Health Service Utilization in Pakistan: Challenging the Policy Makers', *Journal of Public Health*, 27 (1), 49–54.

Soman, Dilip (2010) 'Using Psychology to Encourage Household Savings', Unpublished paper, Rotman School of Management, University of Toronto, 10 March 2010.

Stone, Deborah (1989) 'Causal Stories and the Formation of Policy Agendas', *Political Science Quarterly*, 104 (2), 281–300.

Wade, Robert (1991) *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization* (Princeton University Press).

Wiarda, Howard (1996) *Corporatism and Comparative Politics: The Other Great 'ism'* (Armonk: M.E. Sharpe).

Wong, Joseph (2004a) *Healthy Democracies: Welfare Politics in Taiwan and South Korea* (Ithaca: Cornell University Press).

Wong, Joseph (2004b) 'Democratization and the Left: Comparing East Asia and Latin America', *Comparative Political Studies*, 37 (10), 1213–37.

Wong, Joseph (2011) *Betting on Biotech: Innovation and the Limits of Asia's Developmental State* (Ithaca: Cornell University Press).

World Bank (1993) *The East Asian Miracle: Economic Growth and Public Policy* (New York: Oxford University Press).

9
Methods for the Analysis of Policy Interdependence

Fabrizio Gilardi

Introduction

Policy making takes place in a context of interdependence. This statement is uncontroversial and there are many examples of cases in which decision-makers in one country (or state, city, and so on) are somehow influenced by the choices made in other countries (or states, cities, and so on). For instance, consider the following exchange on 'Big Society', a project by David Cameron, the British Prime Minister, aiming to reinvigorate civil society in the context of drastic cuts in public spending:[1]

> Francis Maude (the minister for the Cabinet office) was unconcerned about the unevenness of services that Big Society is likely to entail. In his view, one of the programme's key virtues was its potential for heterogeneity. 'People will associate to form a bigger, stronger society in many ways which will be random,' he said. 'It'll be fantastically different in different places.' 'What if it's fantastically better in some places?' I asked. 'The advantage of where we are with technology is that it becomes much easier for the ones where it isn't fantastic to look at what's going on where it's fantastic and draw from it,' he replied.

This argument is remindful of the line made famous by US Supreme Court Justice Louis Brandeis in 1932 in defence of the view that decentralization enables policy innovation and the spread of best practices: 'It is one of the happy incidents of the federal system that a single courageous State may, if its citizens choose, serve as a laboratory; and try novel social and economic experiments without risk to the rest of the country.'[2]

Many cases of policy interdependence have received considerable academic attention, and some have even generated their own literatures. For instance, the idea of tax competition is premised upon the fact that the tax rates set in one jurisdiction are shaped by those of other jurisdictions (Genschel and Schwarz, 2011). Second, following Vogel's (1995) influential work, 'California effect' has come to denote configurations in which stricter regulations set by important markets (in Vogel's original work, the automobile emission standards set by California) influence the behaviour of private actors in exporting countries and, possibly, even the policies of the latter. Finally, it is well known that smoking bans spread first within the United States (Shipan and Volden, 2006) and then across European countries and beyond.

Thus, policy interdependence is ubiquitous. At the same time, the methodological consequences of interdependence are seldom taken into account systematically, despite the fact that they correspond to the famous 'Galton's problem' (Ross and Homer, 1976), which is known to affect virtually any comparative analysis. The name derives from a point raised by Francis Galton in response to a presentation at an anthropology conference in 1889. Galton argued that the customs of different societies could originate from a single model, in which case the number of independent observations that can be leveraged in a comparison is, in fact, just one (Tylor, 1889). Of course, the problem affects most comparisons in public policy. For instance, if we analyze the determinants of the adoption of smoking bans in Europe, we need to take into account the fact that European countries were influenced by the American example. In social policy, the fact that researchers categorize welfare states as 'Beveridgean' or 'Bismarckian' is an explicit reminder that many countries constructed their social security systems based on the same model.

In most texts on the comparative method, the fact that units of analysis are not independent from one another is usually mentioned briefly as an unavoidable complication, but it is seldom treated in depth. Moreover, interdependence is more than a nuisance. It is a phenomenon that is interesting in itself with significant implications for policy making, such as the appropriate degree of centralization or harmonization in federal systems, in the European Union, or at the international level. Given this importance, it is not surprising that policy interdependence has been studied from a variety of theoretical perspectives. However, the methodological implications and options for comparative policy analysts are usually not discussed explicitly. By contrast, the goal of this chapter is precisely to show how interdependence can be taken into account systematically.

The chapter is structured as follows. The second section unpacks interdependence by distinguishing three related concepts, namely policy diffusion, policy transfer, and policy convergence. Both diffusion and transfer emphasize interdependence explicitly. They share the basic idea that policy choices in one unit are shaped by policy choices in other units although they put emphasis on different aspects of this idea. By contrast, policy convergence, that is, the increased similarity of policies over time can, but need not, be related to interdependence. However, convergence is often discussed in this context. The third section discusses the quantitative tools that comparative scholars can employ to analyze policy interdependence in its various manifestations, while the fourth section presents qualitative strategies. On the quantitative side, the main methods are spatial regression and dyadic analysis. In the former, interdependence is measured through 'spatial lags', that is, weighted averages of the dependent variable in other units, where weights reflect meaningful forms of interdependence. In the latter, units of analysis are pairs of countries (or states, cities, and so on), which make it possible to integrate directly relational information such as whether two countries are neighbours or share the same language. On the qualitative side, case selection and, especially, within-case analysis and process tracing, allow researchers to uncover detailed evidence on the mechanisms driving diffusion, which is harder to do with quantitative methods. The conclusion of the chapter considers how quantitative and qualitative methods can be combined in the study of policy interdependence.

Conceptual issues

Policy interdependence has been studied from three perspectives relying on different concepts, namely transfer, diffusion, and convergence. The following definitions are commonly accepted. They are very general; regardless of the specific formulation, they apply to any type of policy-making authority.

Policy transfer is 'the process by which knowledge about policies, administrative arrangements, institutions and ideas in one political system (past or present) is used in the development of policies, administrative arrangements, institutions and ideas in another political system' (Dolowitz and Marsh, 2000: p. 5). For instance, Jacobs and Barnett (2000) analyzed the role of a particular taskforce in importing and adapting health policy proposals from the United States and the United Kingdom in New Zealand.

Policy diffusion occurs 'when government policy decisions in a given country are systematically conditioned by prior policy choices made in

other countries' (Simmons et al., 2006: p. 787). For instance, Gilardi et al. (2009) found that countries were more likely to adopt hospital financing reforms if other countries appeared to benefit from them.

Policy convergence means that policies become increasingly similar over time (Bennett, 1991). For example, based on expenditures data, Starke et al. (2008) concluded that OECD welfare states have actually not become significantly more similar since the 1980s and, therefore, there has been only limited convergence.

Transfer and diffusion are clearly very close. Some authors have highlighted their differences (Marsh and Sharman, 2009), but the two concepts refer essentially to the same phenomenon despite the fact that the literatures remain distinct, especially with respect to the methodological approaches that are used (transfer analyses tend to be qualitative, while diffusion analyses are mostly quantitative). By contrast, convergence should not be confused with either transfer or diffusion. While the latter are concerned with how interdependence shapes the policy-making process, convergence refers to its outcome. Importantly, the increase in similarity of policies can, but need not, be a consequence of interdependence. While it can be that policy-makers adopt similar policies because they influence one another, it is entirely possible that convergence is the result of common pressures, either in the form of similar problems calling for similar solutions or because many units are exposed to the influence of powerful organizations or countries. The relationship between the European Union and member states, or between a federal state and subnational units, or the World Bank or the IMF and countries in need of financial support, are all cases in point. Although, strictly speaking, these forms of vertical influence are not part of a strict definition of policy interdependence, we include them in the discussion because many studies consider them an integral part of transfer, diffusion, and/ or convergence.

Table 9.1 compares the three concepts with respect to their focus, the dependent variable, and methodology. Convergence is focused on the outcomes of interdependence, it takes changes in policy similarity as

Table 9.1 Policy convergence, transfer, and diffusion (adapted from Holzinger et al., 2007)

	Convergence	Transfer	Diffusion
Focus	Outcome	Process	Policy change
Dependent var.	Outcome	Adoption/process	Adoption
Method	Quantitative	Qualitative	Quantitative

the dependent variable, and it is usually studied quantitatively. Transfer is focused on the process of change and can take both the adoption of the policy and the process of change as dependent variables. The analysis is usually qualitative. Finally, diffusion is focused on policy change, the dependent variable is the adoption of the policy, and the methodology is mostly quantitative.

A common theme in all three literatures is the quest for mechanisms. That is, scholars do not simply ask whether policy interdependence can be observed empirically. They are usually concerned with the factors that make policies transfer, diffuse, or converge. The terminology varies, but the common idea is that interdependence has different shapes and, therefore, the transfer, diffusion, and convergence of policies are influenced by different factors. For instance, working within the policy transfer tradition, Bennett (1997) found that different types of logic account for the spread of different innovations in bureaucratic accountability. The institution of the ombudsman spread through lesson drawing, freedom of information legislation as a legitimation instrument for policymakers, and data-protection laws as a consequence of harmonization within international organizations. Similarly, Shipan and Volden (2008) analyzed the diffusion of antismoking policies among US cities and concluded that it was driven by four mechanisms, namely, learning from early adopters, competition among cities in the same region, imitation of larger cities, and coercion by state governments. Finally, in their study of the convergence of environmental policies among 24 OECD countries, Holzinger et al. (2008) found evidence that these policies became significantly more similar especially as a result of international harmonization and transnational communication but not because of regulatory competition.

Even this brief overview makes it clear that, while these literatures are all concerned with mechanisms, they do not employ a consistent terminology. However, most mechanisms can be usefully grouped under four broad headings (Gilardi, 2012).

Learning is the process whereby the experience of other units influences policy making because it can be used to estimate the likely consequences of policies. For instance, policy-makers may become more willing to adopt a policy if they see that it enjoyed high popular support in other countries, or if it has proven effective in addressing a given problem.

Competition is the fact that policy-makers anticipate or react to the policies of other units in order to attract or retain resources such as investments. Tax competition is the prototypical example here.

Emulation emphasizes the socially constructed properties of policies such as norms of appropriateness emerging from social interactions. In the long run, some policies may be accepted as the normal response to given issues. For instance, same-sex marriage has become much more accepted in recent years.

Coercion underlines the influence of powerful countries, international organizations, and other actors. IMF conditionality or EU conditionality are cases in point.

Thus, the conclusion of these literatures is that interdependence matters, but it takes different forms. That is, policies spread in an interdependent process that can be driven by different mechanisms. But the question is: How do we know? How can we use empirical information to establish whether interdependence matters, and if it does, why? The next section considers quantitative options, and the section following it, qualitative strategies.

Quantitative approaches

The second section has shown that interdependence is not just a nuisance or complication that needs to be fixed. It is an interesting phenomenon in its own right. However, interdependence is also a methodological issue that, if it is not taken into account, can lead to biased results. Many statistical techniques such as regression assume the independence of observations. If this is not the case and interdependence is not modelled appropriately, then analyses will produce unreliable results (Beck et al., 2006; Franzese and Hays, 2007). There are two main options to model interdependence. The first relies on spatial lags, namely, weighted averages of the dependent variable, which are included in the analysis as an additional variable, while the second reshapes datasets so that units of analysis are pairs of countries (or states, cities, and so on) and relational variables can be included directly. The analysis of convergence does not focus on interdependence itself but, rather, on changing degrees of similarity between units. This can be measured in two ways. First, we can look at how the variation of policies changes over time (σ-convergence). Second, we can examine the relationship between policy levels at the beginning of the observation period and policy change during this same period (β-convergence).

Interdependence
A first set of methods can be employed to measure the existence and nature of interdependence. In principle, these methods are relevant for

both the transfer and diffusion literatures, but, because the former is to a large extent qualitative, they are found especially in the latter.

The first method is spatial regression (Beck et al., 2006; Franzese and Hays, 2007; Ward and Gleditsch, 2008). These models capture interdependence through the so-called spatial lag, which is added to the regression as an additional variable. As mentioned earlier, a spatial lag is simply a weighted average of the dependent variable in other units. If the dependent variable is a policy, as is usually the case, then the spatial lag measures the average policy in other units. The idea is quite simple and, in fact, is often used informally by researchers. To illustrate, suppose that we are interested in the unemployment replacement rates of OECD countries (Allan and Scruggs, 2004). Our model will typically include a series of political, institutional, and economic variables, such as the partisan composition of governments, the number of veto players, and economic openness. In this case, the spatial lag would measure, for each country, the average value of unemployment replacement rate in other countries, weighted following a theoretical criterion.

Figure 9.1 shows a simplified example. The top panel shows the so-called connectivity matrix for five countries, which includes information on their linkages. In this simple example, two countries are considered connected if they share a border, and not connected if they do not. The bottom panel

Connectivity matrix

	DEN	FRA	GER	ITA	SWI
DEN	0	0	1	0	0
FRA	0	0	1	1	1
GER	1	1	0	0	1
ITA	0	1	0	0	1
SWI	0	1	1	1	0

Row-standardized connectivity matrix

Dependent variable

Spatial lag

$$
\begin{pmatrix}
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1/3 & 1/3 & 1/3 \\
1/3 & 1/3 & 0 & 0 & 1/3 \\
0 & 1/2 & 0 & 0 & 1/2 \\
0 & 1/3 & 1/3 & 1/3 & 0
\end{pmatrix}
\begin{pmatrix}
63.8 \\
70.8 \\
65.6 \\
45.2 \\
77.2
\end{pmatrix}
=
\begin{pmatrix}
65.6 \\
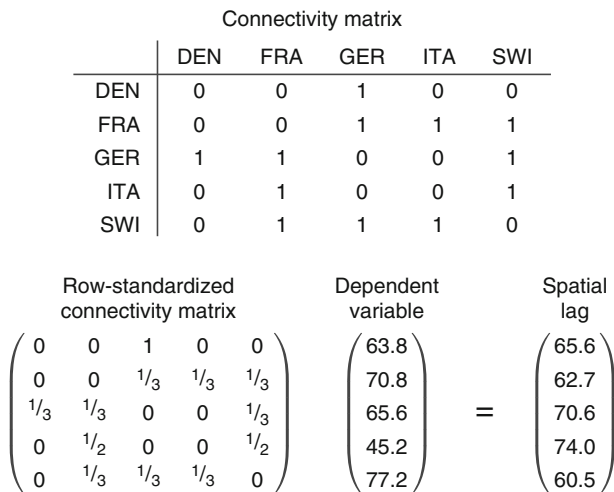62.7 \\
70.6 \\
74.0 \\
60.5
\end{pmatrix}
$$

Figure 9.1 Construction of a spatial lag: Unemployment replacement rates, 2000 (Allan and Scruggs, 2004)

shows how the spatial lag is constructed, namely, by multiplying the connectivity matrix (after it has been row-standardized) with the dependent variable. Taking the example of France, the procedure works as follows:

$$0 * 63.8 + 0 * 70.8 + 1/3 * 65.6 + 1/3 * 45.2 + 1/3 * 77.2 = 62.7$$

That is, the value of the spatial lag for France is equal to the average of replacement rates of its (in this simplified example) three neighbours, Germany, Italy, and Switzerland.

Row-standardization means that each cell is divided by the sum of the corresponding row, which makes it possible to interpret the spatial lag intuitively as a weighted average.[3] Thus, 65.6 is the unemployment replacement rate of Denmark's only neighbour (Germany), 62.7 is the average replacement rate of France's three neighbours (Germany, Italy, Switzerland), and so on. The construction of the connectivity matrix is a crucial step of the analysis, in which theory plays an important role. The connections should be defined such that they can be linked with diffusion mechanisms as clearly as possible. Geography is a useful starting point, but it is usually an ambiguous indicator because it cannot be clearly linked to a specific mechanism. Moreover, geographic neighbourhood may be misleading as some units may be technically neighbours but, in reality, are not strongly related to one another. Therefore, although the task is not easy, researchers should try to come up with the best possible indicators for diffusion mechanisms. For instance, in their study of the adoption of the flat rate tax, Baturo and Gray (2009) measured learning by including in the connectivity matrix information on one relevant outcome of that policy, namely foreign direct investment. Another example is Gilardi and Wasserfallen (2014), which used data on commuter flows across jurisdictions to measure tax competition among Swiss cantons. Commuter flows are a good indicator because if many people work in, say, Zurich but live in Aargau (a neighbouring canton), it means that many people could move from the former to the latter canton to benefit from lower taxes without disrupting their professional lives. Thus, Zurich is likely to be responsive to the tax policy of Aargau. In this example, the indicator overlaps with geography, but it is not always the case. For instance, Ticino and Valais are technically neighbours but, in fact, they are linked only by a mountain road that is closed for the better part of the year. Thus, it is unlikely that these two cantons can compete for taxpayers. However, this is the assumption if we operationalize competition with simple geographic proximity, as is often done.

As Beck et al. (2006) put it, 'space is more than geography', and researchers should be creative and find measures that go beyond

geographic distance and permit measurement of interdependence in a more theoretically meaningful way.

An alternative method is the dyadic approach, in which units of analysis are pairs of countries (or states, cities, and so on). This approach is common in the international relations literature because many of the phenomena it studies are dyadic in nature. Conflict, trade, and treaties, for instance, all occur between countries and cannot be measured if one takes countries as observations. The dyadic approach was adapted to the study of policy interdependence by Volden (2006), who investigated the diffusion of policies within the Children's Health Insurance Program (CHIP) among US states. The units of analysis in this study were all logical pairs of states (except same-state pairs), and the dependent variable was defined as taking the value of one if, in a given year, the first state of the dyad (the potential 'receiver' of the policy) adopts a policy that was already in place in the second state of the dyad (the potential 'sender'). The goal of the analysis, then, is to find out how the probability that this occurs varies as a function of characteristics of the first state, of the second state, and of their relationship.

The advantage of the dyadic approach is that it can easily incorporate many variables at all three levels (sender, receiver, and the relationship between the two), which is much more complicated within a spatial regression framework. Moreover, in the dyadic approach, we can take into account several dimensions of the dependent variable at the same time (for instance, tax rates for different income categories), which is much less straightforward with spatial regressions. Conversely, if the dependent variable has just one dimension, then some of the added value of the dyadic approach is lost.

Using this method, Volden (2006) found evidence of policy learning. States were more likely to adopt a policy that another state already had if the latter was successful, namely, if it could increase insurance rates among poor children, which is one of the main objectives of CHIP. Other examples of studies using the dyadic approach applied to policy diffusion are Gilardi (2010) and Füglister (2011), both of which, like Volden (2006), focus on learning.

There are many methodological issues that one needs to consider when using the dyadic approach (for an overview, see Gilardi and Füglister, 2008). An especially important point is the construction of the dependent variable. Table 9.2 shows a simplified example with four units (A, B, C, and D) and two periods. In this example, only D is shown as the potential sender, but in reality the dataset would comprise all logical combinations of units. The example assumes that the policy has two dimensions; indeed, the dyadic approach is most useful when policies

Table 9.2 An artificial dyadic dataset

Unit_i	Unit_j	t	DV_1	DV_2	DV_3	DV_4	Policy_1i	Policy_2i	Policy_1j	Policy_2j
A	D	1	1		0.36		0.90	0.30	0.70	0.60
A	D	2		1	0.22	−0.14	0.60	0.40	0.70	0.60
B	D	1	1		0.45		0.30	0.40	0.70	0.60
B	D	2		0	0.41	−0.04	0.60	0.20	0.70	0.60
C	D	1			0.21		0.65	0.80	0.70	0.60
C	D	2	0	0	0.34	0.13	0.55	0.90	0.70	0.60

consist of several dimensions because it can incorporate them more easily than spatial regression can. Handling multidimensional policies is one of the added values of the dyadic approach. Thus, the first line in the table refers to the dyad AD, where A is the potential receiver and D the potential sender. 't' denotes the period, the next four columns ($DV_1 - DV_4$) four possible dependent variables, and the last four columns ($Policy_{1,i} - Policy_{2,j}$) the values of two policy dimensions for A and D.

Given this setup, there are several ways in which the dependent variable can be constructed.

The first is to code the dependent variable 1 if A moves closer to D on at least one policy dimension. Following this rule, DV_1 is coded 1 for the dyad AD because both $Policy_{1,A}$ and $Policy_{2,A}$ move closer to D. On the other hand, DV_1 is coded 0 for the dyad CD because C moves away from D on both dimensions. The dyad BD is more ambiguous. $Policy_{1,B}$ moves closer to $Policy_{1,D}$, but there is a movement in the opposite direction for $Policy_2$. Consistent with the rule, DV_1 is coded 1 in this case. However, another reasonable coding rule could be that the dependent variable is coded 1 if the two units become more similar on one dimension without becoming more dissimilar on another. If we follow this criterion, DV_2 is again coded 1 for AD and 0 for CD, but now BD must also be coded 0. As the number of dimensions increases (for instance, there are six in Volden, 2006, and four in Füglister, 2011), coding the dependent variable dichotomously following these rules can quickly become impractical, and decisions are based on ad hoc thresholds. As an alternative, we can compute the Euclidean distance between units in a multi-dimensional policy space as shown in Figure 9.2. The *x*-axis represents the policy position on the first dimension and the *y*-axis that on the second. The Euclidean distance between units is represented by the dashed lines connecting units A, B, and C (both at time 1 and 2) with D, and it can be used to code a third dependent variable (DV_3). Hence, 0.36 is the distance between A and D at time 1. Finally, DV_4 can be coded as the difference between DV_3 at time 1 and at time 2. Interestingly, DV_2 was coded 0 for BD because of the inconsistent direction of change in the two policy dimensions, but DV_4 shows that, on the whole, B has moved closer to D between time 1 and time 2. In sum, the construction of the dependent variable in a dyadic framework is a crucial analytical step for which there is no straightforward solution, but several plausible options.

In conclusion, spatial regression and the dyadic approach are two alternatives to study policy interdependence quantitatively. The former offers a well-understood way to model interdependence in the context
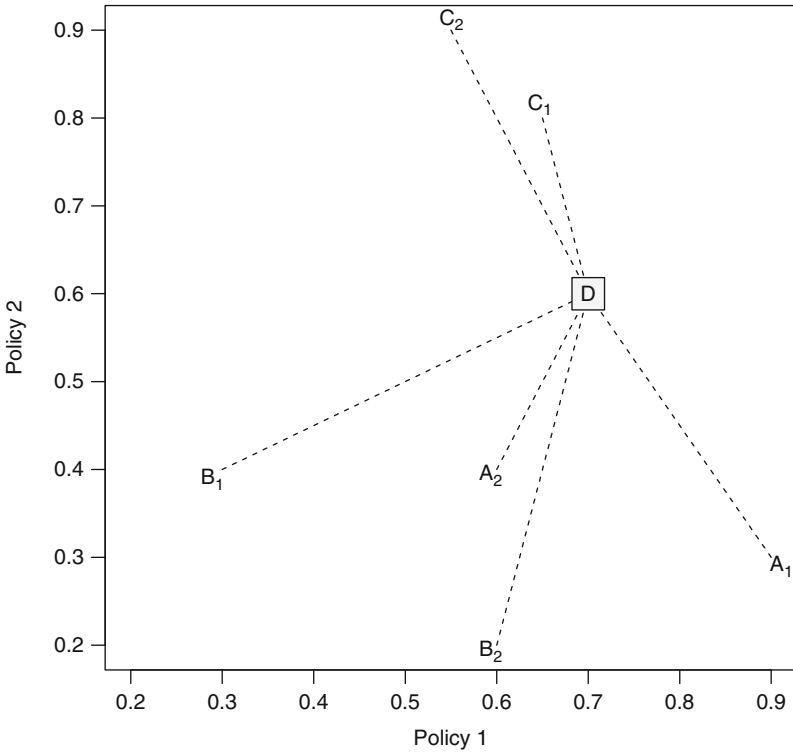
Figure 9.2   Construction of the dependent variable in the dyadic approach

of standard time-series cross-section datasets while the other gives more flexibility, especially when working with several policy dimensions, at the cost of additional methodological complications.

Convergence

There are two standard ways to measure convergence quantitatively, which rely on the concepts of σ- and β-convergence. The terms are technical, but the meaning is intuitive. σ-convergence means simply that we look at the variation of policies across units at different points in time. Convergence occurs if variation decreases over time. β-convergence means that units that are further away from the eventual convergence point have more catching up to do.

The concepts are illustrated in Figures 9.3 and 9.4 using unemployment replacement rates (Allan and Scruggs, 2004). Consistent with the idea that convergence means that policies become more similar over time,
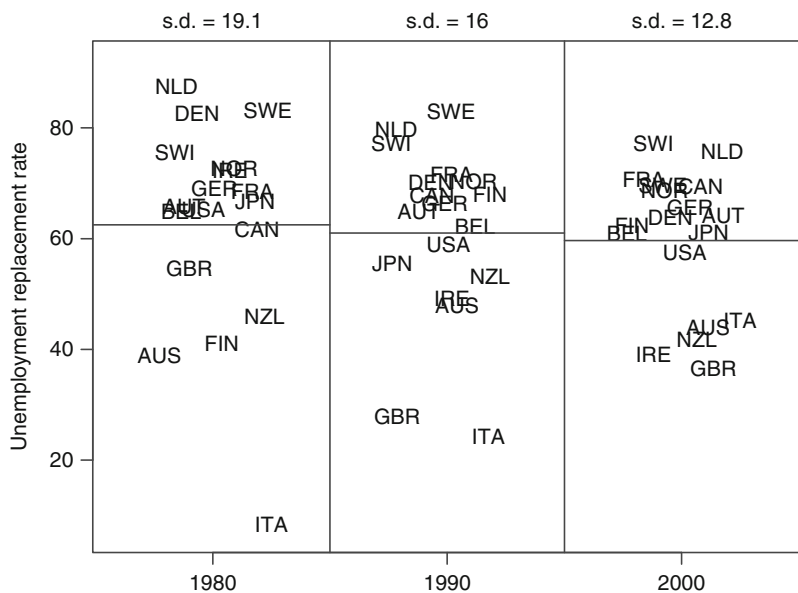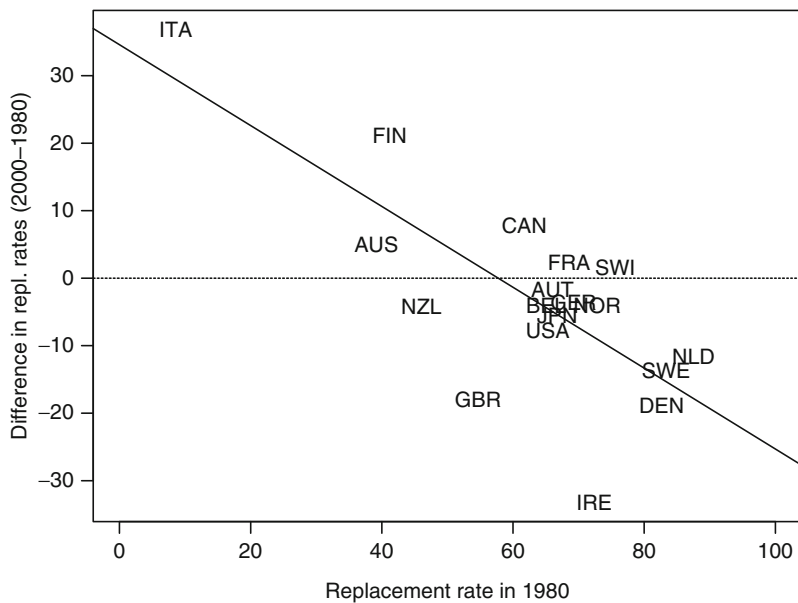
Figure 9.3   σ-convergence

Figure 9.4   β-convergence

σ-convergence is based on a measure of variation such as the standard deviation. Figure 9.3 shows quite clearly that the unemployment replacement rates of OECD countries have become more similar between 1980 and 2000, and the standard deviations shown on the top of the graph clearly demonstrate it. For instance, the spread between the Netherlands and Italy is visibly smaller in 2000 than in 1980. At the same time, the average (shown by the horizontal lines) remained essentially unchanged, so that we have here a good example of convergence to the mean.

β-convergence focuses on another aspect of the phenomenon, namely that, if units are to become more similar, those with more extreme values at the beginning of the observation period must undergo greater adjustment. Thus, we should observe a negative relationship between the initial level of the policy and its change over time, as shown in Figure 9.4. Italy had the lowest replacement rates in 1980 and a greater increase between 1980 and 2000, while in countries such as Denmark, Sweden, and the Netherlands, replacement rates were above average in 1980 but have decreased since then. The formal test is a regression of change on initial values in which we expect a negative and significant coefficient (which, in conventional notation, is called *β*), which is dutifully the case in our example.[4]

These measures look simple enough, but in fact, there are several complications. Plümper and Schneider (2009) note that convergence can easily remain undetected by σ-convergence. For example, if convergence occurs in clusters and the clusters are sufficiently apart from one another, then policy variation in the whole sample can increase even though it decreases within the clusters. For instance, it could be argued that, in Figure 9.3, the standard deviation underestimates the extent of convergence. In effect two clusters are visible: a larger one, in which replacement rates range from 60 to 80 per cent in 2000, and a smaller one, where replacement rates are lower. Within these clusters, the decrease in variation is probably larger than implied by the standard deviation computed for all countries. Similarly, when employing *β*-convergence, researchers should carefully look not only at convergence in the whole sample but also within clusters, conditional on contextual factors, and at different levels. The point here is simply that the methods assume convergence at a global level, but this can occur also within subgroups. If this is not recognized, both σ- and *β*-convergence can produce misleading results.

Furthermore, Holzinger (2006) argues that convergence analysis is more difficult when one considers different levels of details in the measurement of policies such as their presence or absence, the types of instruments applied, and the settings of these instruments. Therefore, she advocates a dyadic approach similar to that discussed earlier. The main difference is

that dyads are undirected, that is, the dependent variable captures similarities but without differentiating between potential senders and potential adopters. An example of application is Holzinger et al. (2008), which investigated 40 environmental policies and 89 policy items. As mentioned earlier, this study found that convergence was driven mainly by international harmonization and transnational communications.

Qualitative approaches

Qualitative approaches to policy interdependence have been used especially in the policy transfer literature, but examples also exist for policy diffusion. In contrast to quantitative approaches, there are few distinct tools to tackle the specific question of interdependence. However, in a recent article, Starke (2013) examined some exemplary works and discussed how critical steps of qualitative analysis, such as case selection (see Chapter 3 in this book) and process tracing (see Chapter 4 in this book), can be employed for the study of policy transfer or diffusion.

First, case selection could benefit from a 'diverse cases' strategy, which 'has as its primary objective the achievement of maximum variance along relevant dimensions' (Gerring, 2007: p. 97). A traditional method of difference could in principle be adopted, namely, by selecting cases with different outcomes, similar control variables, and different diffusion variables. Alternatively, the method of agreement would require that cases differ on the outcome and on key diffusion variables, but are similar on the control variables. However, it is well known that Mill's methods do not work well in practice because cases never fit cleanly in Mill's schemes. Therefore, it is advisable to strive for good variation on diffusion mechanisms instead of trying hopelessly to find 'most similar' or 'most different' cases or to squeeze them acrobatically into these categories. At any rate, case selection alone gives only limited analytical leverage and must always be combined with other strategies.

Second, and most important, qualitative researchers should highlight the nature of interdependence by concentrating on the process of diffusion within cases. This step corresponds to what is known as process tracing. There is no doubt that a fine-grained focus on process and mechanisms is the most important contribution that qualitative work can offer to the understanding of diffusion. Thus, qualitative research should strive to uncover crucial 'causal-process observations', that is, 'an insight or piece of data that provides information about context, process, or mechanism, and that contributes distinctive leverage in causal inference' (Brady and Collier, 2004: p. 277). For instance, Weyland (2007) showed in detail how pension privatization in Chile played an

important role in reforms in other Latin American countries. In Bolivia, a crucial event was the Finance Minister's budget director attending a keynote speech by the architect of Chile's pension privatization; similarly, in El Salvador the Chilean model was put on the agenda through a consultant who was involved in the Chilean reform and who was originally hired to assist with a smaller-scale project (Weyland, 2007: p. 101). In contrast, contacts with experts and policy-makers from Argentina and Colombia, which had also introduced reforms of the pension system, were much more limited (Weyland, 2007: pp. 105–106).

Another example of detailed within-case analysis is Biedenkopf (2011), which studied how EU environmental legislation affected policies in the United States, both at the federal and at the state level. One of the findings is that the influence of the European Union was quite differentiated. For instance, in Washington State, legislators took inspiration from the European Union at the conceptual level rather than for concrete details. At the same time, many actors in this state demonstrated a good knowledge of EU legislation, which leads the author to conclude that learning has taken place, at least to some extent. This was not always the case. In California, for instance, policy-makers strived to achieve a leadership role within the United States by following the example of the European Union, but they paid less attention to the concrete details of EU legislation, which indicates that the emulation mechanism was stronger than learning in this case.

In some instances, researchers may even uncover 'smoking guns' supplying very strong evidence. For example, in his study of national tax blacklists, Sharman (2010) provides examples of countries that literally copied and pasted legislation from others. The most striking case is Venezuela (Sharman, 2010: p. 625):

> [T]he Venezuelan legislation made reference to the wishes of the Mexican legislature and the need to be consistent with the Mexican constitution. Worse still, the original Mexican list had included Venezuela, and thus by copying the Mexican list, Venezuela succeeded in blacklisting itself.

Such 'causal-process observations' are absent from purely quantitative studies but are essential for a good understanding of policy interdependence, and they can also play an important role in theory building. In sum, the strength of qualitative approaches is their capacity to identify processes and mechanisms in detail. Case selection is important but not as central as process tracing for the study of policy interdependence.

Conclusion

Policy interdependence means that the policies of a given unit (country, state, or city) are shaped by those of others. It is a ubiquitous phenomenon that has been studied from different perspectives, notably policy transfer, policy diffusion, and policy convergence. Conceptually, scholars have focused on the nature of interdependence and have formulated different mechanisms that characterize it, such as learning, competition, emulation, and coercion. Methodologically, the quantitative arsenal is well stocked. Convergence can be measured with standard concepts such as σ- and β-convergence, even though several issues need to be taken into account.

Interdependence can be integrated within standard regression models via spatial lags, which are weighted averages of the dependent variable, where weights are defined to capture theoretically meaningful relationships among units. Alternatively, the dyadic approach considers each unit as a potential sender and a potential receiver of a policy and examines what factors make it more likely that the receiver adopts a policy that the sender already has. This strategy allows for considerable flexibility to model various types of interdependencies and to take into account policies that consist of multiple dimensions but at the cost of additional methodological complications. In particular, the construction of the dependent variable is not straightforward and the data structure becomes more complex. Turning to qualitative approaches, there are fewer standard procedures for the analysis of policy interdependence but their contribution is both crucial and unique. In particular, they allow to uncover important 'causal-process observations' that demonstrate the presence of interdependence and help to interpret its nature on the basis of theoretical mechanisms.

It is commonplace to say that research benefits from the integration of qualitative and quantitative methods, but it is true. Despite several methodological contributions (for example, Lieberman, 2005), how exactly this should be done remains unclear. However, the comparative advantages of quantitative and qualitative methods for the study of policy interdependence are obvious. The former offers consolidated procedures to detect interdependence and, to some extent, to uncover some clues on the mechanisms behind it. This is not something that qualitative approaches do well. However, only qualitative research can provide a detailed account of what interdependence means exactly and of the extent to which theoretical mechanisms can be found in reality. The most powerful analyses are certainly those that can combine these two dimensions, no matter if the combination is a little ad hoc or does not fit nicely into a grand methodological scheme.

Notes

1   Lauren Collins. 'All Together Now! What's David Cameron's Big Society about?', *The New Yorker*, 25 October 2010.
2   New State Ice Co. v. Liebmann" (http://goo.gl/CZPmi, accessed 28 November 2012).
3   On the pros and cons of row-standardization, see Plümper and Neumayer (2010).
4   Difference 2000 – 1980 = 34.61(8.72) – 0.60 (0.13) × repl. rate. in 1980 (OLS regression, standard errors in parentheses).

Bibliography

Allan, James P. and Lyle Scruggs (2004) 'Political Partisanship and Welfare State Reform in Advanced Industrial Societies', *American Journal of Political Science*, 48 (3), 496–512.

Baturo, Alexander and Julia Gray (2009) 'Flatliners: Ideology and Rational Learning in the Adoption of the Flat Rate Tax', *European Journal of Political Research*, 48 (1), 130–59.

Beck, Nathaniel, Kristian S. Gleditsch, and Kyle Beardsley (2006) 'Space is More than Geography: Using Spatial Econometrics in the Study of Political Economy', *International Studies Quarterly*, 50 (1), 27–44.

Bennett, Colin J. (1991) 'Review Article: What Is Policy Convergence and What Causes It?', *British Journal of Political Science*, 21 (2), 215–33.

Bennett, Colin J. (1997) 'Understanding Ripple Effects: The Cross-National Adoption of Policy Instruments for Bureaucratic Accountability', *Governance*, 10 (3), 213–33.

Biedenkopf, Katja (2011) 'Policy Recycling? The External Effects of EU Environmental Legislation on the United States', PhD dissertation, Vrije Universiteit Brussel, unpublished manuscript.

Brady, Henry E. and David Collier (eds) (2004) *Rethinking Social Inquiry. Diverse Tools, Shared Standards* (Lanham: Rowman and Littlefield).

Dolowitz, David P. and David Marsh (2000) 'Learning from Abroad: The Role of Policy Transfer in Contemporary Policy-Making', *Governance*, 13 (1), 5–24.

Franzese, Robert J. and Jude C. Hays (2007) 'Spatial Econometric Models of Cross-Sectional Interdependence in Political Science Panel and Time-Series-Cross-Section Data', *Political Analysis*, 15 (2), 140–64.

Füglister, Katharina (2011) 'Where Does Diffusion Take Place? The Role of Intergovernmental Cooperation for Policy Diffusion in Switzerland', *European Journal of Political Research*, 51 (3), 316–49.

Genschel, Philipp and Peter Schwarz (2011) 'Tax Competition: A Literature Review', *Socio Economic Review*, 9 (2), 339–70.

Gerring, John (2007) *Case Study Research: Principles and Practice* (Cambridge University Press).

Gilardi, Fabrizio (2010) 'Who Learns from What in Policy Diffusion Processes?', *American Journal of Political Science*, 54 (3), 650–66.

Gilardi, Fabrizio (2012) 'Transnational Diffusion: Norms, Ideas, and Policies', in Walter Carlsnaes, Thomas Risse, and Beth Simmons (eds) *Handbook of International Relations,* 2nd edn (London: Sage).

Gilardi, Fabrizio and Katharina Füglister (2008) 'Empirical Modeling of Policy Diffusion in Federal States: The Dyadic Approach', *Swiss Political Science Review*, 14 (3), 413–50.

Gilardi, Fabrizio and Fabio Wasserfallen (2012) 'How Socialization Attenuates Tax Competition', *British Journal of Political Science*, forthcoming.

Gilardi, Fabrizio, Katharina Füglister, and Stéphane Luyet (2009) 'Learning From Others: The Diffusion of Hospital Financing Reforms in OECD Countries', *Comparative Political Studies*, 42 (4), 549–73.

Holzinger, Katharina (2006) 'Methodological Pitfalls of Convergence Analysis', *European Union Politics*, 7 (2), 271–287.

Holzinger, Katharina, Christoph Knill, and Thomas Sommerer (2008) 'Environmental Policy Convergence: The Impact of International Harmonization, Transnational Communication, and Regulatory Competition', *International Organization*, 62 (4), 553–87.

Holzinger, Katharina, Helge Jörgens, and Christoph Knill (2007) 'Transfer, Diffusion und Konvergenz: Konzepte und Kausalmechanismen', *Politische Vierteljahresschrift*, Sonderheft 38, 11–35.

Jacobs, Kerry and Pauline Barnett (2000) 'Policy Transfer and Policy Learning: A Study of the 1991 New Zealand Health Service Taskforce', *Governance*, 13 (2), 185–213.

Lieberman, Evan S. (2005) 'Nested Analysis as a Mixed-Method Strategy for Cross-National Research', *American Political Science Review*, 99 (3), 435–52.

Marsh, David and Jason C. Sharman (2009) 'Policy Diffusion and Policy Transfer', *Policy Studies*, 30 (3), 269–88.

Plümper, Thomas and Christina J. Schneider (2009) 'The Computation of Convergence, or: How to Chase a Black Cat in a Dark Room', *Journal of European Public Policy*, 16 (7), 990–1011.

Plümper, Thomas and Eric Neumayer (2010) 'Model Specification in the Analysis of Spatial Dependence', *European Journal of Political Research*, 49 (3), 418–42.

Ross, Marc Howard and Elizabeth Homer (1976) 'Galton's Problem in Cross-National Research', *World Politics*, 29 (1), 1–28.

Sharman, Jason C. (2010) 'Dysfunctional Policy Transfer in National Tax Blacklists', *Governance*, 23 (4), 623–39.

Shipan, Charles R. and Craig Volden (2006) 'Bottom-Up Federalism: The Diffusion of Antismoking Policies from U.S. Cities to States', *American Journal of Political Science*, 50 (4), 825–43.

Shipan, Charles R. and Craig Volden (2008) 'The Mechanisms of Policy Diffusion', *American Journal of Political Science*, 52 (4), 840–57.

Simmons, Beth, Frank Dobbin, and Geoffrey Garrett (2006) 'Introduction: The International Diffusion of Liberalism', *International Organization*, 60 (4), 781–810.

Starke, Peter (2013), 'Qualitative Methods for the Study of Policy Diffusion', *Policy Studies Journal*, 41 (4), 561–82.

Starke, Peter, Herbert Obinger, and Francis G. Castles (2008) 'Convergence Towards Where: In What Ways, If Any, Are Welfare States Becoming More Similar?', *Journal of European Public Policy*, 15 (7), 975–1000.

Tylor, Edward B. (1889) 'On a Method of Investigating the Development of Institutions: Applied to Laws of Marriage and Descent', *Journal of the Anthropological Institute of Great Britain and Ireland*, 18, 245–72.

Vogel, David (1995) *Trading Up: Consumer and Environmental Regulation in a Global Economy* (Cambridge: Harvard University Press).

Volden, Craig (2006) 'States as Policy Laboratories: Emulating Success in the Children's Health Insurance Program', *American Journal of Political Science*, 50 (2), 294–312.

Ward, Michael D. and Kristian S. Gleditsch (2008) *Spatial Regression Models* (Los Angeles: Sage).

Weyland, Kurt (2007) *Bounded Rationality and Policy Diffusion: Social Sector Reform in Latin America* (Princeton University Press).

10
Gendering Comparative Policy Studies: Towards Better Science

Amy G. Mazur and *Season Hoard*

Introduction[1]

Feminist policy scholars in Western Europe first acknowledged the empirical gaps and gender biases in theory and methodology used in the study of the state and policy in the early 1980s. By the early 1990s, researchers in North America and Australia joined their Western European counterparts in the new feminist academic enterprise that sought to systematically study the interconnections between the social construction of men's and women's identities, policy and the state; in other words to 'gender' the study of state action. In the mid 1990s, a loose methodological consensus developed within this transnational community around conventions for conducting research, developing theory and reporting findings; a consensus, which moved the new field, Feminist Comparative Policy (FCP), into a stage of vitality and institutionalization. In 2012, with over 400 published pieces, an estimated 20 million euros in research funding, over 100 active researchers and four journals that serve as publication outlets – *Social Politics*, *International Journal of Feminist Politics, Politics and Gender* and *Women, Politics and Policy* – FCP holds an important place in comparative policy studies and political science more broadly speaking.

Given the theoretical and methodological connections to and commonalities with comparative policy studies and the continuing political relevance of gender policy issues throughout the world, FCP has a great deal to offer the comparative study of public policy. Overall, as this chapter will show, gendering policy analysis fills crucial empirical gaps in our knowledge about the state and policy in postindustrial democracies and, in doing so, places issues of democratic governance and performance to the fore. This process of gendering policy analysis also

entails the development of mixed methodological approaches that emphasize conceptualization and measurement and their application to complex context-sensitive studies that focus on problem-driven analytical puzzles.

One of the most successful efforts of contributing to non-FCP analysis, work that has 'gendered' the study of the welfare state, illustrates well the importance of bringing gender in.[2] Beginning with a critique of Esping-Andersen's typology of welfare state regimes (1990), feminists asserted that crucial dimensions of the comparative taxonomy were missing: women's access to paid work; the nature of women's and men's roles within the family; and the extent to which the social policies of a given country allowed women to combine work and family without being economically dependent on men. Subsequent work developed new gendered typologies and conducted comparative analysis with them that brought gender in as a category of analysis. Today, most comparative scholarship incorporates this crucial gendered dimension. Indeed, not taking into consideration these gender issues means that critical aspects of social rights and inclusion, issues at the core of the welfare state research agenda on postindustrial democracies, are excluded from the analysis. Thus, the process of gendering welfare state research enhanced knowledge and theorizing about the welfare state and social policy more generally speaking.

Unfortunately, the case of comparative welfare state research tends to be an exception; scholars in 'mainstream' or non-feminist policy studies tend to ignore this burgeoning area. The goal of this chapter is to show the methodological contributions that FCP scholarship makes to comparative policy studies; contributions which, like the gendering welfare state example, have the potential to strengthen comparative work on public policy from the larger theory-building and research design perspective of this volume. Thus, a central argument of this chapter is that non-feminist policy scholars need to better integrate FCP study designs, methods and findings in order to make comparative policy studies more systematic and, ultimately, to do better science.

The first section of the chapter presents FCP as an area of study since it first emerged in the late 1980s, its core features and more recent methodological contributions and issues. The successful gendering welfare state case is used as a running example to illustrate these core features. Next, the four major streams of FCP research – gendering welfare states, feminist policy formation, women's movements and policy and state feminism – as well as three new trends – representation, intersectionality and feminist institutionalism – are presented in terms of how they gender comparative policy analysis and the added value they bring to conducting

policy research from a comparative perspective. The final section provides two illustrative examples of research designs in two of the research streams – state feminism and feminist policy formation – to further show the dynamics and contributions of doing gendered research in comparative policy analysis. The chapter concludes with a discussion of the methodological lessons learned from FCP research.[3]

FCP as an area of study

FCP assesses how, why and to what end the contemporary state has responded to demands for the advancement of women's rights, gender equality and for striking down gender based inequities in society. In other words, to what degree has the state become feminist?

Core features[4]

Feature 1 – Using Gender as a Category of Analysis: Since the mid 1980s, feminist research across different disciplines has shifted its focus from sex, a more or less dichotomous variable based on biological differences between men and women and/or women alone, to gender; the social construction of sexual difference between men and women. As Joan Scott first asserted in 1986, the relational concept of gender should be the prime 'category of analysis' in theoretical frameworks and research designs. This holistic approach to the use of gender is intended to push analyses beyond the 'add women and stir' phase where sex or women is added as an analytical afterthought. As many others have pointed out, a feminist approach to the study of politics, 'feminist political science', necessarily places gender at the center of analysis 'as a relational concept and an analytic category …' (Krook and Mackay, 2011: p. 4), within the context of the promotion of feminist action and ideas.[5]

Following this broader trend in feminist studies, FCP work has placed gender as a fundamental category of analysis in a variety of ways as the analysis of the major streams of research shows below. Indeed, the central question of FCP – about whether states are feminist – is inherently gendered. These include: if and how public policy strikes down social hierarchies based on the asymmetrical relations between men and women and social constructions of those identities and/or patriarchy; if and how gender-specific actors, individual women, women's movements and groups, gender experts, men and many others influence or block state action; and if and how state structures have incorporated approaches that include perspectives and demands that address gender relations and asymmetries.

Gendering Welfare States Example: The complex notion of gender is brought into the feminist conceptualization of welfare states. It goes beyond a focus on women alone to how the division of labour between men's and women's roles in the family impacts the very nature of the welfare state and how market economies and welfare state 'de-commodify' women's work and not men's. Welfare state studies had either completely overlooked any complex consideration of gender (for example, Esping-Andersen, 1990) – a 'gender-blind' approach – or had included sex-based variables, for example, women's labour force participation, as indicators of social policy impact without placing that dimension into a larger context of complex gender relations; in other words, the 'add women and stir' approach (for example, Schmidt, 1993).

Feature 2 – An Integrative Feminist Empirical Approach: FCP scholars have increasingly followed an 'integrative empirical feminist (Mazur, 2011)' approach to studying gender, policy and the state. As Figure 10.1 shows, more generally in the context of feminist epistemologies, an empirical integrative approach combines empirical, postmodern and standpoint approaches, the three major feminist epistemologies identified by philosophers of science (for example, Harding, 1986).

FCP scholars conduct studies of political phenomena to contribute to theory that also seeks to solve larger social problems and is methodologically pluralist, replete with the 'creative tensions' (Siim, 2004) found in much feminist political science work. Often, members of the FCP community design studies that are 'problem-driven' or 'use-inspired'
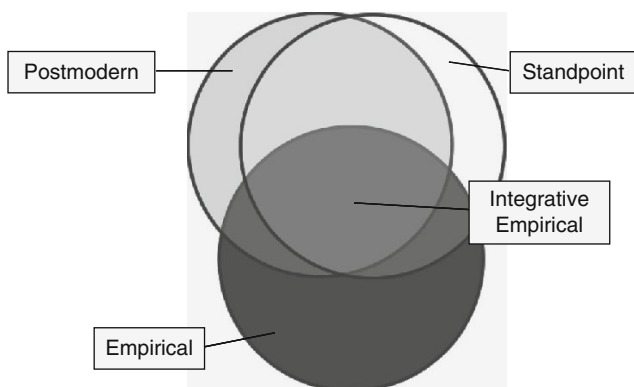

Figure 10.1   Mapping an integrative empirical feminist approach

(Stokes, 1997) – a major trend across all social sciences – so that findings may be used to help policy practitioners and activists in women's policy agencies, political parties, movements and organizations learn more about the causes of gender-based inequities and the complex range of solutions, including different ways of designing good practices. Since the early 1980s, FCP scholars have been consulted regularly in their expert capacity by the European Union – in particular, in the context of the Action Plans on gender equality – the Council of Europe, the Nordic Council, the United Nations, the World Bank as well as numerous country-based commissions at national and subnational levels. At the same time that FCP researchers have had an eye towards impact evaluation studies, they also focus on theory-building issues regarding the major drivers of policy dynamics or the 'ingredients' for gender policy successes.

> *Gendering Welfare States Example*: Clearly, the gendering welfare state scholarship reflects the goal of developing new gendered conceptualizations and theoretical frameworks to advance empirical analysis; the welfare state regime taxonomies were gendered to allow for more accurate analysis of what welfare states were actually doing. The creative tension has been clear. While the original impulse was to develop a feminist critique of the treatment of the welfare state, many of the analysts moved quickly to developing new frameworks and taxonomies, built from the non-feminist work that had ignored gender as a complex dimension.[6] The feminist welfare state literature is decidedly problem driven. How can welfare states in democracies mediate between the need to regulate the market and develop women-friendly and gender-sensitive policies that promote gender equality? In this area as well, we have seen qualitative approaches applied to reconceptualizing and quantitative large-N analyses of empirical findings, often within the same studies, for example, O'Connor et al. (1999). Many of the experts of gender and welfare states have become policy experts for the European Union and other international organizations. A noteworthy example is Orloff's analysis of the gendering welfare state literature for the United Nations (2002).

Feature 3 – Operationalizing Feminist Theory on Democracy: Normative political theory has been an integral part of the development of feminist studies in general. A major question asked by the large literature on feminist theory is whether established Western democracies are as democratic as observers think, particularly given the degree to which women and women's issues have been excluded from politics in the context of

the formal articulation of universal and gender-blind values of equality, freedom and representation.[7] Feminist theorists who write on democracy have argued for a better inclusion of women and ideas that favour women's rights in the political process through Pitkin's (1967) typology of representation. Two of Pitkin's major categories have been adopted for analysis: 'descriptive' representation, the presence of women in public office and 'substantive' representation, the inclusion of women's interests in a meaningful manner in public discussions and policy. These themes of representation and democracy are crucial to FCP studies in so far as they ask the empirical question of whether, how and why democratic states can be feminist, identifying the presence of women as a potential key variable for policy change. The question is less about the specific form and design of democracy than its capacity to incorporate women's interests and women's representatives as a formerly excluded group into the political process and, in doing so, to promote gender equality and a more complete democratic system.[8]

> *Gendering Welfare States Example*: While the gendering welfare states scholarship did not directly address political representation as much as the other areas of FCP that are further examined below, the reconceptualization of the welfare state was rooted in normative feminist theories that argue social policies of postindustrial democracies should include feminist considerations of gender-based inequities in market, family relations and state policy in order to meet the basic prerequisites of any democratic system.

Feature 4 – Bringing the Patriarchal State Back in as a Research Question: – As most political scientists agree, particularly since the state was 'brought back in' in the 1980s (Skocpol, 1985), the concept of the state – government structures as opposed to country – is not a simple idea. For many feminist theorists, the state is highly problematic given that it is a product of systems of power based on male-domination or patriarchy. From the assumption of the patriarchal nature of the state, where state actions, structures and actors seek to perpetuate systems of gender domination that keep women in their inferior positions in the public and private spheres and men as key players in maintaining that domination, many feminist analysts have dismissed or have been highly critical of the state as an arena for positive social change (for example, Mackinnon, 1989). Other feminist theorists provide a more malleable view of state patriarchy and argue that certain state arenas may be appropriate sites for feminist action (for example, Pringle and Watson, 1992). FCP analysts do not

entirely dismiss the possibility of a patriarchal state; they see the issue of state patriarchy as a question for empirical research. Some parts of the state may be patriarchal while other parts may have the potential to be quite women-friendly. FCP has placed the state and its institutions and the roles men and women play in those processes at its analytical core.

> *Gendering Welfare States Example*: The gendered taxonomies of welfare state regimes are based on the assumption that states may eschew uniquely patriarchal regimes. The most feminist/women-friendly category of welfare states, usually associated with the Nordic countries, is seen to have policies that emphasize family models based on men and women sharing parenting roles and getting equal access to the labour market; lower categories of countries develop policies that prop-up more established gender norms about female full-time caregivers and male breadwinners with relatively lower levels of full-time women's labour force participation.

Feature 5 – Comparative Theory-Building in Western Postindustrial Democracies: FCP scholars have used principles of research design and methods developed outside of a feminist perspective to pursue a comparative theory-building agenda. Up until 2000, FCP work utilized small-to medium-N analysis – case studies and the comparative method – and took a 'most similar systems approach', where economic and political development in Western postindustrial democracies are the control variables and variations in nation-based culture, state-society relations, women's movement mobilization, government design and so on are examined as they influence gender, state and policy issues. Most FCP work assumed that postindustrial countries from the West, unlike other countries of the world, shared certain contexts and institutions with notable cross-national variations. A part of the common heritage is that women's movements have developed strategies aimed, at least in part, to influence the democratic policy process and the development of large welfare states. Designing and using concepts, like feminism, that are able to 'travel' across national boundaries without 'stretching' the core meaning has also been a key part of the comparative agenda of FCP (Sartori, 1970).

> *Gendering Welfare State Example*: The development of new gender comparative classification systems of welfare states is based on the notion of a most similar systems design, where the contemporary postindustrial state has reached common levels of commodification

and social policy regulation to manage market economies with large service sectors; gender role change was a part of that level of postindustrial development. Indeed, the feminist critique came out of the empirical realization that gender norms were changing and that women's movements were indirectly a vector of that change. Although Japan has been included in some of the gender welfare state analyses, for the most part these analyses have focussed on Western postindustrial democracies.

Feature 6 – The Glass Wall between Feminist and Non-Feminist Political Science: Many FCP scholars since the mid-1990s have actively sought to intersect their work with non-feminist literature. Instead of completely rejecting traditional political studies or uncritically using feminist studies, FCP work often purposefully develops the strengths and shores up the weaknesses of each to advance knowledge in both areas. In general, efforts to intersect work with non-feminist political science tend to be one-way and appear to confront an 'opaque glass wall' that prevents feminist research from being used in a meaningful way by scholars who do not take a feminist frame (Mazur, 2011).

> *Gendering Welfare States Example and Feminist Policy Formation as a Counterpoint*: The gendering welfare states example is the outlier in terms of actually changing the study of welfare states. As Pierson's review article in *Comparative Political Studies* shows, gender has become an important dimension of welfare state studies (2000). An early example of the glass wall can be found in Outshoorn's use of non-feminist policy literature of the time in her analyses of abortion and equal employment policies in the Netherlands (Outshoorn, 1986, 1991), including Lowi's policy typology (1964), Bachrach and Baratz's non-decision-making theories (1970), Schattschneider's notion of problem definition (1960) and Cobb and Elder's work on agenda setting (1983). In fact, much of the feminist policy formation work in FCP has used these core notions as important launching points. In subsequent developments of the rich non-feminist research on preadoption (for example, Kingdon, 2003; Baumgartner et al., 2006), these applications to feminist policy have been rarely, if ever mentioned or referenced. A recent special issue of the *Revue Française de Science Politique* on agenda setting, problem definition and feminist policy formation (Boussaguet and Jacquot, 2009) indicates the degree to which the comparative policy literature continues to ignore insights from feminist work on preadoption issues.

Recent developments: Methodological contributions and issues

We now turn to the methodological contributions and issues in FCP that have emerged more recently through the assessment of nine large-scale FCP projects carried out since 2000, presented in Table 10.1 (Mazur, 2009).

In some ways, gender and comparative policy scholarship has not changed in recent years. Gender remains a fundamental category for analysis; issues of patriarchy, the persistence of gender-biased norms and the state are at the center of study designs; feminist and non-feminist theory continues to be operationalized in studies, and comparative theory building based on qualitative analysis is an important part of research designs. At the same time, there have been significant new developments reflected in these projects, also identified by other feminist policy scholars (for example, Squires, 2007). While these methodological issues define the shape of current and future policy scholarship that takes a gendered perspective, they also resonate with the methodological agenda of comparative policy studies covered in this book.

Table 10.1   List of FCP research projects

Women's Movements and Reconfigured States (1997–2003) Banaszak et al. (2003)

Gendering Europeanization (1999–2003) Liebert (2003)

RNGS (1995–2009). Research Network on Gender Politics and the State http://libarts.wsu.edu/polisci/rngs (for example, McBride and Mazur, 2010)

EGG (2002–2005). Enlargement, Gender and Governance: The Civic and Political Participation and Representation of Women in Central and Eastern Europe http://www.qub.ac.uk/egg/ (for example, Galligan et al., 2007)

MAGEEQ (2003–2007). **Policy Frames and Implementation Problems: The Case of Gender Mainstreaming** http://www.mageeq.net/ (for example, Lombardo et al., 2009)

QUING (2006–2011). Quality in Gender Equality + Policies http://www.quing.eu/ (for example, Verloo and Walby, 2012)

FEMCIT (2006–2010). Gendered Citizenship in Multi-Cultural Europe: The Impact of Contemporary Women's Movements http://www.femcit.org (Halsaa et al., 2012)

VEIL (2006–2009). Values Equality and Differences in Liberal Democracies http://www.univie.ac.at/veil/

Governing Difference (2006–2009). A Challenge for New Democracies in Central and South Eastern European Countries

---

The norm of international research groups: Towards a cosmopolitan scientific community?

FCP practitioners began to develop international research networks in the early 1980s and in the 1990s created multinational research projects, securing significant funding to maintain formal research groups with publications, meetings, websites and newsletters. The networks have often met at the conferences of the European Consortium of Political Research, the American Political Science Association, the International Political Science Association or the International Studies Association.[9] Since 2000, while not a prerequisite, the creation of a large international research group has become more of the norm; all but one of the nine projects presented in Table 10.1 had formal international research groups. The FCP community tends to be quite global and 'cosmopolitan', rather than driven by national 'parochial' considerations.[10] Most groups have both European and North American members with the leadership not being dominated by a single nationality. The FCP community more broadly speaking is quite evenly distributed between the Western countries, when controlling for population. In 2012, there were roughly one hundred scholars who worked regularly on FCP research.

Going global: How big of an 'N'?

The inclusion of East Central and South Eastern European countries into the analytical purview of many of the newer FCP projects opened the door for the consideration of countries from outside of the West with different levels of economic and political development and cultural contexts. Moreover, the analytical perspective of FCP has clearly gone beyond the nation-state to include a multilevel approach, where the sub-national and extra-national levels are just as, if not more, important than the national level. This push to go beyond the West clearly resonates with the calls made by other feminist political scientists for a 'Comparative Politics of Gender' that places the study of gender and politics into a systematic cross-national and cross-regional perspective.[11] Indeed, much of the current research in FCP has sought broader and more systematic comparison through the inclusion of countries from different regions of the world.

From a quantitative perspective these newer studies have the potential to increase the number of observations or N, to boost their explanatory power and to bring in considerations of most 'different' systems. At the same time, expanding the number of countries in a given study to include countries outside of the postindustrial West raises a new set of analytical issues that stem out of the inclusion of radically different

cultural and political contexts. Thus, researchers must carefully assess whether their concepts can actually 'travel' across these divergent settings without being 'stretched' beyond their original meanings (Sartori, 1970); a consideration that has been an integral part of FCP scholarship since its beginnings.

Methodological pluralism: Good or bad?

Reflecting FCP's integrative feminist empirical approach, methodological pluralism has become a more pronounced attribute of research in this area. To be sure, much work is qualitative, emphasizing the importance of expert analyses of country cases, process tracing, elite interviews and archival research. Contrary to conventional wisdom about feminist studies, FCP researchers also use quantitative large-N analysis, sometimes out of necessity, due to the tendency to include more countries in study designs (for example, Kittelson, 2008; Bolzendahl, 2009; Huber et al., 2009; Htun and Weldon, 2012). In addition, the tools of medium-N analysis, Qualitative Comparative Analysis (QCA) (Ragin and Rihoux, 2008) are being increasingly used in FCP studies (for example, Mazur, 2009; McBride and Mazur, 2010). Moreover, many studies are seeking to follow a mixed methods approach (O'Connor et al., 1999; Htun and Weldon, 2010; McBride and Mazur, 2010; Weldon, 2011; Hoard, 2014; Montoya, 2013).An emerging part of the feminist approach is to more formally conceptualize and to develop specific data collection and analysis techniques, drawing from both feminist and non-feminist work; many make the call for more systematization and formalization of concepts and measurements. Developing valid and reliable concepts for empirical analysis has been an important theme in many studies and an arena for fruitful collaboration between feminist and non-feminist analysts (for example, Goertz and Mazur, 2008).

More recent feminist research has embraced the European 'discursive turn' (Kantola, 2006).[12] These studies have brought in approaches based on feminist standpoint theory and social constructivism with a focus on framing, discourse and policy content. The MAGEEQ/QUING projects developed a general approach to analyzing the inclusion of gender and other areas of inequalities called Critical Frame Analysis (CFA). These more interpretivist scholars may reject formalization of concepts and measurements and the scientific method more generally speaking; thus there is a potential for fundamental epistemological disagreements between FCP scholars that may make project completion challenging.

Therefore, the shift towards methodological pluralism is not necessarily positive. On one hand, the key issues of whether formal polices are

effectively implemented to actually change gender relations in society is left relatively unexamined due to the focus on discourse and policy content; on the other hand, an absence of clearly articulated hypotheses, formal concepts and findings may limit the broader empirical and applied messages of the studies. Without this systematic consideration of policy implementation and impact, the feminist scholarship that focuses on discourse, frames and the 'positionality' of researchers in the process and the power system, risks to miss the central questions of FCP – whether all of this new state activity actually matters in terms of actually achieving a more gender equal society. At the same time, the increased emphasis on bringing research to public officials and citizens through public meetings, conferences and training, an emphasis brought in by more constructivist approaches, potentially outweighs the empirical gaps.

Representation

Normative and empirical questions of democracy are now a focal point for FCP studies in terms of including women's movements, actors and interests into the democratic process to make stable democracies more democratic and for newer democracies in terms of effective transitions to democracies. Women's movements and women's policy agencies, and not just legislatures, are now identified as potential sites of representation as well (Weldon, 2002a). This emphasis on representation as a means to link issues of women's presence to policy outcomes in the FCP scholarship dovetails with comparative scholarship on women in politics, which is also taking a more systematic look at women's substantive representation and policy outcomes (for example, Celis and Childs, 2008; Celis and Mazur, 2012).

Big theory to answer big questions

While the 'critical tensions (Siim, 2004)' between the different strands within the integrative empirical approach of FCP may lead to some differences about the appropriate methodological tools for data collection and analysis, there seems to be an emerging consensus around the development of more macro-level theory to answer the 'big questions' of democracy, gender equality and representation.[13] The inclusion of countries from throughout the world, at 'all levels and in all domains' in study designs, allows for an analytical terrain on which a variety of propositions and hypotheses can be tested about what the major drivers are for achieving maximum democratic performance and for promoting successful and meaningful gender equality policies. The development of culturally sensitive analytical concepts that travel across time and space

as well as increasing the number of observations also lends itself to the eventual pursuit of more macro theory that can apply to a larger number of cases. These studies, thus, can provide more informed and more scientific answers to the perennial big questions being asked by FCP, questions which are at the center of Comparative Policy Studies as well as all areas of political science.

What matters the most: The path towards configurational analysis and complexity

Big theories do not necessarily mean mono-causal explanations. Indeed, research on FCP, often conducted at the mid-range level,[14] has posited a series of key factors that must be taken into consideration as hypotheses/ potential drivers in producing authoritative policy responses in any macro theory of gender, policy and the state. These include ideas (for example, Orloff and Palier, 2009; Sauer, 2010), institutions (for example, Krook and Mackay, 2011), women's movements (for example, McBride and Mazur, 2010; Outshoorn, 2010; Weldon, 2011; Htun and Weldon, 2012) and partisan politics (for example, Kittelson, 2008). Increasingly, studies have been identifying sectoral differences to be more important in explaining policy outcomes than national or regional policy styles (for example, Mazur, 2009; McBride and Mazur, 2010; Htun and Weldon, 2012). While individual studies make claims to the explanatory power of specific factors and often give a privileged place for these as drivers and objects of analysis, many others show that pathways to feminist success in policy are actually quite complex, often being comprised of different 'configurations' of factors and in some cases with different pathways leading to the same outcome (Banaszak et al., 2003; Krook, 2009; Mazur, 2009; McBride and Mazur, 2010; Weldon, 2011). This analytical logic resonates with qualitative methodologies and with the tenants of equifinality and conjunctural causation that are part and parcel of Comparative Configurational Analysis as a general methodological approach (Rihoux and Ragin, 2008 and their chapter in this volume).

Persistence of the glass wall

One characteristic of FCP which has remained through the years is the degree to which non-feminist policy studies and political science continue to ignore gender and policy research; thus the glass wall persists. With the exception of welfare state research, mainstream comparative politics and policy studies still do not integrate the findings of feminist scholarship in a meaningful way or bring gender, women's movements or women's representation as an important aspect to be analyzed in

comparative studies of democratic politics. To be sure, there has been an increase in publications of gender research in non-feminist journals and some qualitative methodological discussions have integrated some of the lessons learned from FCP studies (for example, Geortz and Mazur, 2008; Lieberman, 2009), but few non-feminist scholars have seriously gendered their own analysis beyond isolated cases or the more expedient approach of 'add women and stir'.

Research streams of FCP: Gendering non-FCP studies

We now turn to identifying the major streams of current FCP research. Here, we present some of the most recent work from each stream; given the vitality of this area some of the work we cover here has not yet been published or is the object of doctoral research.

Gendering Welfare States – As we have already noted, taking the welfare state broadly construed as the primary object of analysis, the work in this area has arguably had the most success in breaking down the glass wall with non-feminist political science. Taking on Esping-Andersen's taxonomy of the welfare states in Western postindustrial democracy, feminist analysts asserted that any understanding of the contemporary welfare state had to be gendered (1993). In Orloff's influential critique of the non-feminist concept of welfare states, there are two new necessarily gendered dimensions:

> Thus, the decommodification dimension must be supplemented with a new analytic dimension that taps into the extent to which states promote or discourage women's paid employment – the right to be commodified. I call this fourth dimension of welfare-state regimes access to paid work. If decommodification is important because it frees wage earners from the compulsion of participating in the mar-ket, a parallel dimension is needed to indicate the ability of those who do most of the domestic and caring work – almost all women – to form and maintain autonomous households, that is, to survive and support their children without having to marry to gain access to breadwinners' income.
>
> (Orloff, 1993: pp. 318–19)

The concepts of welfare regimes and welfare state taxonomies were thus gendered and applied to welfare state analyses from both feminist and non-feminist perspectives. A recent special issue of *Social Politics,* edited by Palier – a non-feminist political scientist – and Orloff, exemplifies the

way feminist and non-feminist research on welfare state has become integrated (2009). Feminist analysts recently have begun to 'gender' the Varieties of Capitalism framework as well (Mandel and Shalev, 2009). Another distinctive feature of the welfare state stream is that it takes seriously the analysis of impacts. Indeed, the extent to which gendered social policies have actually contributed to changes in women's status often in comparison to men's is at the center of the cross-national (Daley, 2011) and often quantitative analyses (for example, Huber et al., 2009; Bolzendahl, 2009).[15]

Value addition: This first stream of research shows clearly how bringing gender into established classification systems and theoretical frameworks as well as empirical findings can enhance the leverage and analytical power of those theories to produce valid results that reflect the full range of political and social activity that are actually occurring. Without this feminist impulse to gender research in this area, both in terms of frameworks and empirical analysis, welfare regime analysis would only be presenting a partial picture of the dynamics and determinants of welfare state regimes in post-industrial democracies.

Feminist Policy Formation scholarship scrutinizes the ways in which public policy promotes women's status and strikes down gender hierarchies through the study of the obstacles, actors, content and processes of policy that is purposefully feminist. Feminist or gender equality policy is conceptualized as a distinct sector of government action that has a range of subsectors that promote feminist goals across all the areas of government action that have the potential to change gender relations: for example, blueprint, political representation, equal employment, reconciling work and family, family law, anti-sexuality and violence policy and reproductive rights policy.[16] It is important to note that analyses of feminist policy assess the degree to which government action takes a feminist approach or not; thus, a normative assumption here, based on feminist theory, is that the state needs to promote a feminist agenda in these areas.

The state in this area, as in other streams, is conceptualized as a disaggregated entity, where certain parts can pursue feminist activity and other arenas can be positively anti-feminist. The task of policy analysts is to study whether, how, when and why state action actually follows a feminist course whether it is formally articulated in policy documents and statements or not. Given the presence of the feminist impulse in the public arena since the late 1960s, any systematic understanding of the contemporary state, therefore, must take into consideration feminist policy as a separate sector, in its full complexity; in other words, how state action has been specifically gendered.

This subarea of scholarship continues to be very active, with a wide range of feminist subsectors of policy under scrutiny, from an increasingly cross-regional/national perspective. Anti-sexual violence policy has been a recent focus (Weldon, 2002b; Montoya, 2013). A recent special issue of *Comparative European Politics* covers some of the key areas of feminist policy (Mazur and Pollock, 2009a). Sauer (2009) examines veiling policy, a relatively new area of feminist policy in South Eastern European countries. Morgan (2009) conducts a cross-national analysis of the content of reconciliation policies in Western Europe; Engeli (2009) on abortion and artificial reproductive technology policies in eight Western European countries; Zippel (2009) on sexual harassment policy at the EU level; and Haffner-Burton and Pollock (2009) on EU gender mainstreaming. French speaking FCP experts have turned to specifically looking at policies that focus on the private sphere like violence and prostitution policies (Engeli et al., 2008). Lombardo and Forest (2012) continue the tradition of Liebert (2003) and others to focus on the 'Europeanization' of feminist policy.

Like much non-feminist policy scholarship, research in this area takes a policy stage approach. Problem definition has been a major focus of much FCP work, particularly scholarship that takes a discursive/social constructivist approach. Inspired by E.E. Schattschneider's caveat: '… the definition of politics is the supreme instrument of power', this work examines the framing of policy problems and their solutions (for example, Lombardo et al., 2009). A recent special issue of *Revue Française de Science Politique* (Boussaguet and Jacquot, 2009) brought together francophone FCP experts to specifically look at how women's movements contribute to setting the agenda and defining problems on feminist policies in Francophone Europe and the EU level.

Another FCP effort in this stream is seeking to 'gender' the Comparative Agendas Project (CAP) – a large-scale cross-national quantitative project that has coded tens of thousands of policy agendas across most Western European countries, the United States and Canada in 225 plus areas of policy over the past 100 years with the ultimate goal of conducting cross-national analyses on agenda setting.[17] Gender was virtually ignored in the original design of the project – feminist policy was only treated in a single residual category combined with civil rights and minority issues, and there was little effort to determine whether feminist issues were placed on institutional agendas. Annesley et al. (2011) recodes the CAP data in the Netherlands, Great Britain, Switzerland, Denmark and Spain to analyze agenda setting on gender equality. A current collaborative project by the original CAP team and feminist researchers led by Engeli

and Mazur is underway to go through the entire CAP database and recode all policy cases that have the potential to be gendered (94 out of the original 225) for mention of gender and feminist ideas. A pretest on 4000 cases in the US congressional hearings dataset indicates that 2.4 per cent of all gender potential issues actually end up formally mentioning gender.

This focus on agenda setting, problem definition and policy framing highlights the extent to which FCP work has been centred on content, policy adoption processes and outputs rather than outcomes and impact. This is an area of weakness of FCP that undermines feminist policy research to systematically determine whether the state has successfully pursued feminist policy in terms of achieving gender equality in society. While calls for more impact and implementation research have been made (for example, Mazur and Pollock, 2009b), systematic comparative research on the later stages of the policy process is only in its nascent stages and still tends to focus on outputs in implementation rather than impact (for example, Avdeyeva, 2009; Haffner-Burton and Pollock, 2009).

Value addition: Here we see the importance of refocusing the analysis on a broad swathe of policy that had been previously ignored by non-feminist scholarship. The systematic analysis of feminist and gender-specific policy formation in terms of the process, outcomes and impacts of that policy *and* the major causal drivers that promote feminist action provide a more complete picture of state action that reflects the highly complex, variegated and, often, contradictory nature of state action as a whole. Moreover, theorizing state action in terms of feminist action and issues of inclusion of women's interests, provides an additional critical test for democratic performance on a more normative level: Can the contemporary democratic state respond to demands for social change and justice and, if so, to what end and why? As in the case of gendering the CAP, research findings have the potential to provide definitive evidence about the extent to which states have taken on gendered and feminist issues and, hence, provide new data for studying systematically the complex causal factors that will allow us to refine our theories about public policy and the state more broadly speaking.

Women's Movements and Policy research is concerned with the interplay between women's movements, the state and policy. A major issue of interest here is to evaluate the success of women's movements in influencing public policy. Researchers of the women's movement turned to the state and public policy in the 1990s given the degree to which women's movements had sought to engage with the state over the past forty years at all levels – local, subnational and extranational. For feminist analysts

and theorists alike, women's movements are defined as a major potential vector of feminist change and women's representation.

Banaszak et al. (2003) set the stage for this area of FCP research by conducting a comparative study of how women's movements and organizations affected and were affected by 'state reconfigurations' in the 1990s. The mixed methods study of Weldon (2011) asserts that movements, women's and class-based movements, are central to the promotion of democracy, and many other studies have shown that women's movements 'matter' in the development of meaningful state action. FCP work has gone far in conceptualizing and operationalizing what is meant by the women's movement (McBride and Mazur, 2008; Outshoorn, 2010), a crucial step for conducting sound comparative theory-building studies (Mazur et al., forthcoming). Scholars show that contemporary women's movements are significantly different from other social movements in terms of their structure and activities. Women's movements include both autonomous protest-oriented groups and more formal organizations. Similarly, women's movements have developed close relationships with state actors and have sought to promote reform as well as take more protest-oriented stances, typically associating with 'new social movements'. Finally, feminist social movement researchers have shown that contrary to the conventional wisdom that the new era of new social movements has come to a close, women's movements are still alive and active, albeit in different ways.

Value addition: In this third stream, FCP scholars place women's movements on the research agenda as an important variable, if not a crucial determinant in policy formation, that needs to be carefully examined. Women's movement ideas are an important analytical touchstone for studying the content of feminist policy as well. The conceptualization of women's movements beyond merely protest and anti-system groups also pushes non-feminist analysts to examine women's movements separately from other new social movements and suggests that other social movements need to be considered in terms of a more institutionalized and permanent form.

State Feminism research arguably takes the most direct route to understanding how the contemporary state is gendered. It scrutinizes the state structures and actors that are formally charged with promoting women's rights and striking down gender hierarchies: women's policy agencies/gender equality machineries and the agents who work for them – 'femocrats'. As McBride and Mazur (2007) assert, the concept of state feminism went from a 'loose notion' about women's policy agencies to a more precise analytical construct about whether women's policy

agencies worked with women's movements to promote women's interests in the state through both promoting women's movement actors participation and the inclusion of women's movement ideas in the state policy discussions and policy. Thus, women's policy agencies are potentially important vectors of representation and enhanced democratization, they can be crucial arenas for both women's substantive and descriptive representation. Examined in further detail in the next section, the Research Network on Gender Politics and the State (RNGS), a 40 member international research group, has been a major contributor to the research agenda on state feminism since its creation in 1995.

Value addition: A focus on state feminism allows researchers to assess directly the extent to which the contemporary state has seriously taken on demands for social change and justice, in this case, through gender equality, thus once again contributing to solving the larger question of to what degree are consolidated and stable democracies actually democratic. Systematic analysis of women's policy agencies also fills an empirical gap about the state more generally, particularly given all Western postindustrial democracies and all countries of the world, for that matter, have had some form of women's policy office at various levels of governments.

Additional streams or new cross-cutting themes?

The following three areas have more recently become important objects of study and research foci for FCP analysts. Given the extent to which these issues are being taken-up to varying degrees in all current feminist policy work, it is not clear whether each is a discrete stream of research. Nonetheless, they provide new and formerly unchartered avenues for comparative policy analysis to be gendered.

Representation

Issues of linkages between women's descriptive and substantive representation have become central analytical foci for many FCP scholars alongside other political scientists who use Pitkin's taxonomy of representation (Celis and Mazur, 2012). The question of whether women make a difference in public policy formation and whether policies are representing women's interests are at the fore of studies on feminist policy formation. At the same time, women's movements and state feminism research that focus on representation assert that the ultimate test of democratic performance and representation is whether the substantive content of public policy is taking on formerly excluded interests (for example, Celis and Childs, 2008). A methodologically rich, cross-national

and cross-regional literature has emerged on the development and impact of quota policies throughout the world as well (Dahelrup, 2006; and Krook, 2009). This comparative work combines three out of the four FCP streams, given that quota policies are a specific example of political representation policies – one of the feminist policy subsectors – and that women's movements and women's policy agencies are potentially major partners in the adoption and implementation of quotas. Also, the new scholarship on quotas examines not just the content and diffusion of quota policies but also whether they make a difference in enhancing women's substantive and descriptive representation in the crucial dynamics of implementation and impact.

Intersectionality

The relatively new concept of intersectionality, the notion that systems of gender discrimination are interwoven with other systems of discrimination and inequality based on ethnicity, race, class, culture, religion and sexual orientation, is becoming an essential analytical tool as it relates to representation, democracy and gender equality.[18] For example, working-class Muslim women from North Africa have interests different from upper-class white women in European society. Thus, the representation and the equality policies that are formulated to respond to each group's demands may be quite different (Hankivsky and Cormier, 2011).

Many of the most recent FCP studies address intersectionality in their research designs in large part, given that this has been placed on policy agendas in European countries and at the EU level in recent years as well as on the agendas of certain women's movements. Many of the major international feminist policy research projects have brought in some level of intersectionality as well, for example, FEMCIT and QUING. Several recent publications have investigated how intersectionality has been integrated into genderequality efforts in Western Europe. Verloo and Walby (2012) bring together QUING members to reflect upon how the European Union and EU member states have pursued policies that combine different forms of discrimination. Similarly, Kirzsan et al. (2012) conduct a cross-national and regional analysis of government agencies that have formally institutionalized intersectional approaches that will contribute to state feminism scholarship. Halsaa et al. (2012) examine intersectionality in Europe in the context of issues of citizenship. The exact contribution and findings of this new body of work remains to be seen. However as Hankivsky and Cormier (2011) point out, the precise shape and design of the intersectional approach is in flux. Moreover, the potential threats of diluting gender equality efforts in the development of diversity and intersectional policy must be assessed as well.

Feminist institutionalism

The 'new institutionalism' with its emphasis on formal structures, rules and norms is an important touchstone for FCP. Many FCP studies have directly dialogued with the various forms of new institutionalism in their theory building. An international network , Feminism and Institutionalism International Network (FIIN), has been recently created 'to explore the interplay between feminist approaches to gendered institutions and new institutional theory' as its major goal (http://www.femfiin.com/). The group brings together many FCP members from across the globe and has called for and developed a feminist institutional approach to gender analysis.[19] New scholarship on 'gender, politics and state architecture' that focuses on the territorial dimensions of government that address gender policy issues can also be seen as a part of feminist institutionalism (for example, Chappell and Vickers, 2011). A broad focus on gender and institutions cuts across many areas of political analysis and not just policy or state focused analysis; thus feminist institutionalism may not be as much an emerging distinct research stream as FCP work on representation and intersectionality.

FCP research designs in action: State feminism and feminist policy formation

The following two projects capture a range of FCP studies across the different areas of research. The first is a completed study conducted by a large international network with many publications and outputs focused on state feminism, and the second is an in-progress study of a new object of analysis for the feminist policy formation stream, gender expertise. It was originally conducted in the context of a PhD thesis by a single researcher. We show these two cases to illustrate the different stages of FCP projects and to emphasize the importance of doctoral research in the FCP research cycle as well as of large-scale well funded studies conducted by international networks.

RNGS: Gendering theories of policy change, social movements, representation and institutions

Following the lead of Scandinavian, Dutch and Australian feminist scholars, RNGS was created in 1995 to pursue systematically a cross-national study of the interface between women's policy agencies and women's movements in 17 Western postindustrial democracies. Through complex mixed-methods design, which 'chose not to choose' (Mazur and Parry, 1998) between a uniquely qualitative or quantitative approach from the start, over 40 researchers conducted research to produce 5 issue

books, a data base of 130 policy debates with information on 75 variables and a mixed methods analysis of state feminism across five issue areas (McBride and Mazur, 2010).[20]

The RNGS study 'gendered' comparative policy analysis by posing the question of whether, how and why women's policy agencies make states more democratic through promoting the presence of women's movements, actors and ideas in the policy process. The analysis assessed women's representation in individual policy debates across five policy areas with the potential to shift gender relations significantly – abortion, job training, political representation prostitution and an issue of national significance centered on state reconfiguration or 'hot issue'. Researchers conducted qualitative analyses in three policy debates for each country across all five policy areas through process-tracing to determine whether the agencies introduced gendered frames into the policy debates on these issues, whether they supported women's movement claims, whether women's movement actors participated in the debate and whether the policies that were adopted at the end of the debates contained movement demands. The core concepts were formally operationalized and indicators, both qualitative and quantitative, were developed to provide valid and reliable comparative analyses. The study design also was constructed around three strands of theory, both feminist and non-feminist, on representation, movements, policy framing and institution.

Value addition: Thus the multi-methods empirical analysis tested hypotheses from these bodies of theory and the final 'theory of state feminism' produced by that analysis contributed to the cumulative knowledge and theory building in the three strands of theory informing the study. The final multi-methods capstone book in particular made contributions to each area of theory building and, hence, to a more general understanding of democratic performance at the center of comparative policy studies (McBride and Mazur, 2010).

The politics of expertise in postindustrial democracies: Making gender visible and gender expertise matter

Recent non-feminist research has rediscovered the role of expertise in democracy as an important theme in policy studies (Maasen and Weingart, 2005). Yet, this research omits examination of gender expertise. This is a significant oversight given the political relevance of gender policies and the number of organizations, including both non-governmental and governmental, that have sought gender experts to create and inform public policy. Feminist research has found that gender expertise can be

important to feminist policy success (for example, Elgstrom, 2000; Sauer, 2010); however, systematic analysis has yet to be developed.

This study seeks to analyze gender policy expertise and contribute to literature on democratization and representation by analyzing if, when and how women's interests are being conveyed to political actors through gender experts (Hoard,2014). More specifically, it contributes to scholarship on women's substantive representation by analyzing if, how and why the advice and recommendations of gender experts are successfully translated into public policy. Gender experts are therefore conceptualized in terms of their potential to speak for women as a group.

In order to analyze the impact of gender experts, it is necessary to devote considerable attention to the conceptualization and operationalization of gender expertise. This process involves combing insights from several non-feminist policy theories, including Multiple Streams Theory, Advocacy Coalitions Framework, Punctuated Equilibrium Theory, the Theory of Epistemic Communities and Social Construction Theory and integrating this knowledge with current feminist research on gender expertise, gender knowledge, feminist policy formation and women's movements in order to create an operational definition of gender expertise.

The study design utilizes a sequential mixed-methods approach. Semi-structured interviews were conducted with gender experts working with local, regional, national and supranational organizations. These experts come from a variety of backgrounds and performed a variety of activities. Interview data was used to further conceptualize and operationalize the concept of gender expertise as well as to determine factors that bolster or hinder gender expert success in the policy process. Following analysis of the interview data, a survey was developed in order to increase the number and types of experts analyzed. The survey was designed to build on the interview data by testing the factors identified by interview participants as important to their success. In an effort to develop a more macro-level theory, this study also empirically tests the factors that are identified by experts and officials in both the interviews and survey as important for successful implementation of gender expertise through a multilevel, cross-national analysis, which examines gender expertise in public policies of Western postindustrial democracies. QCA is used in order to determine the necessary and sufficient conditions that impact gender expert's success in influencing policy design. The mixed-methods, multilevel and cross-national analysis allows this research to make contributions in theory building to both feminist and non-feminist policy analysis, increasing our understanding of democracy and representation, which are central to both.

Value addition: This research makes gender experts and gender expertise the exclusive focus of its analysis and places gender experts as an important independent variable to be considered when examining public policy, helping to contribute to feminist and non-feminist policy theories through conceptualization and analysis designed to develop better theory regarding expertise. A focus on gender experts and gender expertise allows for assessing the extent to which Western postindustrial democracies and the international community has seriously pursued efforts to achieve gender equality as well as for evaluating the extent to which gender experts are influencing public policy, representing women's interests and the long-term development of more democratic democracies.

Conclusion

Feminist policy scholars have long lamented the gender bias and empirical gaps that exist in non-feminist state and policy research. As this chapter has illustrated, feminist researchers have countered some of the limitations of comparative public policy research through the development of a transnational community of scholars committed to improving feminist and non-feminist research through innovative methodological contributions that strengthen theory building, research design and comparative research in public policy. With an ever-growing international presence in academia and more applied policy settings, FCP is proving to be an important and meaningful area of study within comparative policy studies. As such, there may very well be real scientific ramifications for being uninformed about its contributions.

FCP provides several methodological lessons that can make comparative policy studies more systematic. First, the development of a large and genuinely international research community allows FCP to broaden its impact and contributions through research that goes beyond national analysis and culminate insights through research that is truly multinational. While the majority of this community is comprised of Western scholars, efforts are currently underway to incorporate non-Western gender policy experts and research into newer FCP projects, leading to an area of study that is increasingly going global and expanding its scope beyond postindustrial democracies.

Second, the methodological pluralism and increasing use of mixed methods design result in research that provides more analytical leverage and stronger theories. Third, and especially important to comparative policy

studies, is the emerging emphasis on conceptualization, operationalization and the creation of concepts that travel, a development that also produces more valid and reliable results. Fourth, while tensions exist within FCP between different empirical strands, FCP research increasingly incorporates more comprehensive research designs in order to develop macrolevel as well as mid-range theories about policy content, policy formation and, more recently, implementation and impact. The pursuit of macro-level theory does not mean that FCP scholars are beholden to mono-causal explanations of political phenomena; research in this area illustrates that the determinants of feminist policy success are complex and quite varied. Indeed, the mixed methods designs being used by recent studies provide for better ways of fully grasping *and* theorizing about these highly complex processes.

Fifth, FCP scholarship has contributed to public policy research through the analysis of new objects of study and variables that have previously been ignored in non-feminist literatures such as women's movements, women's policy agencies, gender expertise, feminist policy, gendering policy frames, intersections between race, class, gender, sexuality and so on and the influence of women as actors. Lastly, is the contribution that FCP scholarship has made to feminist and non-feminist theory building through the integration of both literatures. This has led FCP scholars to reconceptualize some of the most central and important concepts and theories about the state, democracy and the policy process in order to conduct research that adequately incorporates women and gender.

In conclusion, FCP scholarship offers several methodological and theoretical insights that can improve comparative policy studies; yet much non-feminist policy research continues to miss the contributions of this important area of study. Given the strong methodological and theoretical links between FCP and comparative policy studies as well as the growing and continued focus on gender policy issues across the globe, the glass wall that still appears to be in place needs to be dismantled. FCP offers much to both feminist and non-feminist policy scholars alike, which, if genuinely incorporated, will lead to a comparative public policy field that is more systematic, integrative, inclusive and, ultimately, more scientific. In the final analysis, therefore, strengthening comparative policy studies through the gendering process will allow this area of political science to better contribute to our general understanding of democratic performance and its critical processes and in doing so enhance democracy itself.

Notes

1 We would like to thank the editors for their valuable and insightful comments as well as the following scholars for sharing with us their feedback and in-progress scholarship: Mieke Verloo, Judith Squires, Celeste Montoya, Andrea Kriszan, Katalin Fabian, Alba Alonso, Emanuela Lombardo, Laurel Weldon, Joyce Outshoorn and Birgit Sauer.
2 See, for example, Orloff (1993) and Sainsbury (1994) for an overview of scholarship in this area.
3 We use here a broad notion of methodology, which recognizes the connections between epistemology, theory building, research design, conceptualization and tools for data collection and analysis.
4 This analysis is based on a large-scale assessment of over 400 FCP pieces (Mazur, 2002).
5 See Lovenduski (1998) for one of the foundational texts on the importance of gender as an analytical concept.
6 See the chapter contributions to Sainsbury's (1994) path-breaking edited volume on 'gendering welfare states'.
7 See Squires (1999) for an overview of feminist theorizing on democracy, gender and the state.
8 For a systematic treatment of the notion 'women's interests' see the 'Critical Perspectives' on Women's Interests in *Politics and Gender* 2011, 7(3).
9 At least 15 FCP books have come out of work carried out at these conferences.
10 Norris (1997)and others have been quite critical of European political science for being overly concerned by parochial and nationalistic issues that have prevented the creation of a unified and systematic European political science.
11 For more on the Comparative Politics of Gender, see the special issue of *Perspectives* (March 2010).
12 See, for example, Lombardo et al. (2009), Verloo and Walby (2012), and Kantola (2010).
13 A workshop that brought together scholars working on these big comparative studies was held at the European Consortium of Political Research joint sessions in April 2012.
14 Following from Merton's (1949) notion of mid-range or meso-theory, mid-range theory building can take place across categories of countries or sectors of policy, for example. Hypotheses from these studies can then be assessed in larger, more macro-level studies across different groups of countries or different sectors of policy.
15 Many of the pieces in *Social Politics* are in the welfare state stream as well.
16 This is the taxonomy of feminist policy developed by Mazur (2002) from the review of the FCP literature in 2000. Blueprint policy includes any policy that seeks to set-up general principles of gender equality like the Equal Rights Amendment in the USA or gender mainstreaming policy, often identified with the European Union (e.g., Haffner-Burton and Pollock, 2009).
17 For more on the CAP go to http://www.comparativeagendas.org/
18 For more on intersectionality as an analytical concept, see, for example, Weldon (2008).

19   For more on feminist institutionalism, see Krook and MacKay (2011).
20   For more on the RNGS study design, its publication and dataset go to http://libarts.wsu.edu/pppa/rngs/

Bibliography

Annesley, Claire, Isabelle Engeli and Francesca Gains (2011) 'The Pace and Profile of Gender Policy Change in Europe', Paper prepared for delivery at the 2011 American Political Science Association Meeting, Seattle, September 2011.

Avdeyeva, Olga (2009) 'Enlarging the Club: When Do Candidate States Enforce Gender Equality Laws?', *Comparative European Politic*s, 7 (1), 158–77.

Bachrach, Peter and Morton S. Baratz (1970) *Power and Poverty. Theory and Practice* (New York: Oxford University Press).

Banaszak, Lee Ann, Karen Beckwith and Dieter Rucht (eds) (2003) *Women's Movements Facing the Reconfigured State* (Cambridge University Press).

Baumgartner, Frank R., Christian Green-Pedersen and Bryan D. Jones (2006) 'Comparative Studies of Policy Agendas', *Journal of European Public Policy*, 13 (7), 959–74.

Bolzendahl, Catherine (2009). 'Making the Implicit Explicit: Gender Influences on Social Spending in Twelve Industrialized Democracies, 1980–99', *Social Politics*, 16 (1), 40–81.

Boussaguet, Laurie and Sophie Jacquot (2009) 'Mobilisations féministes et mises à l'agenda de nouveaux problèmes publics', *Revue française de science politique*, 2 (59), 173–81.

Celis, Karen and Amy G. Mazur (eds) (2012) 'Critical Perspectives: Hanna Pitkin's "Concept of Representation" Revisited: A New Agenda for Studying Representation Rights?', *Politics and Gender*, 8 (4), 508–47.

Celis, Karen and Sarah Childs (eds) (2008) 'Special Issue: The Substantive Representation of Women', *Representation*, 44 (2), 93–204.

Chappell, Louise and Jill Vickers (eds) (2011) 'Critical Perspectives: Gender, Politics, and State Architecture', *Politics and Gender*, 7 (2), 251–93.

Cobb, Roger and Charles D. Elder (1983) *Participation in American Politics: The Dynamics of Agenda-Setting* (Baltimore: Johns Hopkins University Press).

Dahelrup, Drude (ed.) (2006) *Women, Quotas and Politics* (New York: Routledge).

Daley, Mary (2011) 'What Adult Worker Model: A Critical Look at Recent Social Policy Reform in Europe from a Gender and Family Perspective', *Social Politics*, 18 (1), 1–23.

Elgstrom, Ole (2000) 'Norm Negotiations. The Construction of New Norms Regarding Gender and Development in EU Foreign Aid Policy', *Journal of European Public Policy*, 7 (3), 457–76.

Engeli, Isabelle (2009) 'The Challenges of Abortion and Assisted Reproductive Technologies Policies in Europe', *Comparative European Politic*s, 7 (1), 56–74.

Engeli, Isabelle, Thanh-Huyen Ballmer-Cao and Pierre Muller (2008) *Les politiques du genre* (Paris: L'Harmattan).

Esping-Andersen, Gøsta (1990) *The Three Worlds of Welfare Capitalism* (Cambridge: Polity Press).

Galligan, Y., S. Clavero and M. Calloni (2007) *Gender, Politics and Democracy* (Opladen and Farmington Hills: Barbara Budrich).

Goertz, Gary and Amy G. Mazur (eds) (2008) *Politics, Gender and Concepts* (Cambridge University Press).

Haffner-Burton, Emilie and Mark A. Pollock (2009) 'Mainstreaming Gender in the European Union: Getting the Incentives Right', *Comparative European Politic*s, 7 (1), 114–38.

Halsaa, Beatrice, Sasha Roseneil and Sevil Sumer (eds) (2012) *Remaking Citizenship in Multicultural Europe* (Houndmills: Palgrave Macmillan).

Hankivsky, Olena and Renee Cormier (2011) 'Intersectionality and Public Policy: Some Lessons from Existing Models', *Political Research Quarterly*, 64 (1), 217–29.

Harding, Sandra (1986) *The Science Question in Feminism* (Ithaca: Cornell University Press).

Hoard, Season (2014) *Does Gender Expertise Matter? Toward a Theory of Policy Success* (Basingstoke: Palgrave Macmillan).

Htun, Mala and S. Laurel Weldon (2010) 'When Do Governments Promote Women's Rights? A Framework for a Comparative Analysis of Sexual Equality', *Perspectives on Politics*, 8 (1), 207–16.

Htun, Mala and S. Laurel Weldon (2012) 'The Civic Origins of Progressive Policy Change: Combating Violence against Women in Global Perspective, 1975–2005', *American Political Science Review*, 106 (3) 548–69.

Huber, Evelyne, John D. Stephens, David Bradley and François Nielsen (2009) 'The Politics of Women's Economic Independence', *Social Politics*, 16 (1), 1–39.

Kantola, Johanna (2006) *Feminists Theorize the State* (New York: Palgrave Macmillan).

Kantola, Johanna (2010) *Gender and the European Union* (New York: Palgrave Macmillan).

Kingdon, J. (2003) *Agendas, Alternatives and Public Policies* (New York: Harper Collins).

Kirzsan, Andrea, Hege Skeije and Judith Squires (eds) (2012) *Institutionalising Intersectionality: Comparative European Analyses* (Houndmills: Palgrave Macmillan).

Kittelson, Miki Caul (2008) 'Representing Women: The Adoption of Family Leave in Comparative Perspective', *The Journal of Politics*, 70 (2), 323–34.

Krook, Mona Lena (2009) *Quotas for Women in Politics: Gender and Candidate Selection Reform World Wide* (Oxford/New York: Oxford University Press).

Krook, Mona Lena and Fiona MacKay (eds) (2011) *Gender, Politics and Institutions: Towards a Feminist Institutionalism* (Houndmills: Palgrave Macmillan).

Lieberman, Evan (2009) 'Bridging the Qualitative-Quantitative Divide: Best Practice in the Development of Historically Oriented Replication Databases', *Annual Review of Political Science*, 13, 37–59.

Liebert, U. (ed.) (2003) *Europeanisation and Gender Equality. Reframing Public Policy in EU Member States* (Bruxelles: Peter Lang).

Lombardo, Emanuela and Maxime Forest (eds) (2012) *The Europeanization of Gender Equality Policies: A Discursive Sociological Approach* (Houndmills: Palgrave Macmillan).

Lombardo, Emanuela, Petra Meier and Mike Verloo (eds) (2009) *The Discursive Politics of Gender Equality: Stretching, Bending and Policymaking* (Taylor & Francis, e-book).

Lovenduski, Joni (1998) 'Gendering Research in Political Science', *Annual Review of Political Science*, 1, 333–56.

Lowi, Theodore (1964) 'American Business, Public Policy, Case Studies and Political Science', *World Politics*, 16 (4), 677–715.

Maasen, Sabine and Peter Weingart (eds) (2005) *Democratization of Expertise? Exploring Novel Forms of Scientific Advice in Political Decision-Making* (Dordrecht: Springer).

MacKinnon, Catharine (1989) *Toward a Feminist Theory of the State* (Cambridge: Harvard University Press).

Mandel, Hadas and Michael Shalev (2009) 'Gender, Class, and Varieties of Capitalism', *Social Politics*, 16 (2), 161–81.

Mazur, Amy G. (2002) *Theorizing Feminist Policy* (Oxford University Press).

Mazur, Amy G. (2009) 'Comparative Gender and Policy Projects in Europe: Current Trends in Theory, Method and Research', *Comparative European Politics*, 9 (1), 12–36.

Mazur, Amy G. (2011) 'A Feminist Empirical and Integrative Approach in Political Science: Breaking-Down the Glass Wall?', in Harold Kincaid (ed.) *OUP Handbook on Epistemology and Methodology in the Social Sciences* (Oxford University Press), 553–58.

Mazur, Amy G. and Janine Parry (1998) 'Choosing Not to Choose in Comparative Policy Research Design: The Case of the Research Network on Gender, Politics and the State', *Journal of Policy Studies*, 26 (3), 384–95.

Mazur, Amy G. and Mark A. Pollock (eds) (2009a) 'Special Issue: Gender and Public Policy in Europe', *Comparative European Politics*, 7 (1), 1–177

Mazur, Amy G. and Mark A. Pollock (2009b) 'Gender and Public Policy in Europe: An Introduction', *Comparative European Politics*, 7 (1), 1–11.

McBride, Dorothy and Amy G. Mazur (2007) 'State Feminism Since the 1980s: From Loose Notion to Operationalized Concept', *Politics and Gender*, 3 (4), 501–12.

McBride, Dorothy and Amy G. Mazur (2008) 'Women's Movements, Feminism and Feminist Movements', in Gary Goertz and Amy G. Mazur (eds) *Politics, Gender and Concepts: Theory and Methodology* (Cambridge University Press), 219–43.

McBride, Dorothy and Amy G. Mazur (2010) *The Politics of State Feminism: Innovation in Comparative Research*. With participation of Joni Lovenduski, Joyce Outshoorn, Birgit Sauer and Marila Guadagnini (Philadelphia: Temple University Press).

Merton, R. K. (1968 [1949]) *Social Theory and Social Structure* (New York: Free Press) Reprint.

Montoya, Celeste (2013) *From Global to Grassroots: The European Union, Transnational Advocacy, and Combating Violence against Women* (New York: Oxford University Press).

Morgan, Kimberly (2009) 'Caring Time Policies in Western Europe: Trends and Implications', *Comparative European Politics*, 7 (1), 37–55.

Norris, Pippa (1997) 'Towards a More Cosmopolitan Political Science?', *European Journal of Political Research*, 31 (1/2), 17–34.

O'Connor, Julia, Ann Shola Orloff and Sheila Shaver (1999) *States, Markets, Families: Gender Liberalism and Social Policy in Australia, Canada, Great Britain and the United States* (Cambridge University Press).

Orloff, Ann Shola (1993) 'Gender and the Social Rights of Citizenship: The Comparative Analysis of State Policies and Gender Relations', *American Sociological Review*, 58 (3), 303–28.

Orloff, Ann Shola (2002) 'Women's Employment and Welfare Regimes: Globalization, Export Orientation and Social Policy in North America', *Social Policy and Development Paper*, Program Paper 12, United Nations Research Institute for Social Development.

Orloff, Ann Shola and Bruno Palier (2009) 'The Power of Gender Perspectives: Feminist Influence on Policy Paradigms, Social Science and Social Politics', *Social Politics*, 16 (4), 405–12.

Outshoorn, Joyce (1986) 'The Feminist Movement and Abortion Policy in the Netherlands', in Dahrlerup, D. (ed.) *The New Politics of the Women's Movement* (London: Sage), 64–84.

Outshoorn, Joyce (1991) 'Is This What We Wanted? Positive Action as Issue Perversion', in E. Meehan and S. Sevenhuijsen (eds), *Equality Politics and Gender* (London: Sage), 104–21.

Outshoorn, Joyce (2010) 'Social Movements and Women's Movements', in Dorothy McBride and Amy G. Mazur (eds) *The Politics of State Feminism: Innovation in Comparative Research* (Philadelphia: Temple University Press), 143–63.

Pierson, Paul (2000) 'The Three Worlds of Welfare State Research', *Comparative Political Studies*, 33 (67), 791–821.

Pitkin, Hannah F. (1967) *The Concept of Representation* (Berkeley: University of California Press).

Pringle, Rosemary and S. Watson (1992) 'Women's Interests and the Post Structuralist State' in M. Barrett and A. Phillips (eds) *Destabilizing Theory: Contemporary Feminist Debates* (Cambridge: Polity Press), 53–73.

Rihoux, B. and C. C. Ragin (eds) (2008) *Configurational Comparative Methods. Qualitative Comparative Analysis (QCA) and Related Techniques* (Applied Social Research Methods Series, Thousand Oaks: Sage).

Sainsbury, Diane (ed.) (1994) *Gendering Welfare States* (London: Sage).

Sartori, Giovanni (1970) 'Concept Misformation in Comparative Politics', *American Political Science Review*, 64 (4), 1033–53.

Sauer, Birgit (2009) 'Headscarf Regimes in Europe: Diversity Policies at the Intersection of Gender, Culture and Religion', *Comparative European Politics*, 7 (1), 75–94.

Sauer, Birgit (2010) 'Framing and Gendering', in Dorothy McBride and Amy G. Mazur (eds) *The Politics of State Feminism: Innovation in Comparative Research* (Philadelphia: Temple University Press), 193–216.

Schattschneider, E. E. (1960) *The Semisovereign People: A Realist's View of Democracy in America* (New York: Holt, Rinehart & Winston).

Schmidt, Manfred G. (1993) 'Gendered Labour Force Participation', in F. G. Castles (ed.) *Families of Nations* (Aldershot: Dartmouth).

Scott, Joan (1986) 'Gender: A Useful Category of Historical Analysis', *American Historical Review*, 91(5), 1053–75.

Siim, Birte (2004) 'Towards a Contextual and Gender Sensitive European Political Science?', *European Political Science*, 3 (2), 97–101.

Skocpol, Theda (1985) 'Bringing the State Back In: Strategies of Analysis in Current Research', in Peter B. Evans, Dietrich Rueschemeyer and Theda Skocpol (eds) *Bringing the State Back In* (Cambridge University Press), 3–37.

Squires, Judith (1999) *Gender in Political Theory* (Cambridge: Polity Press).

Squires, Judith (2007) *The New Politics of Gender Equality* (Houndmills: Palgrave Macmillan).

Stokes, Donald E. (1997) *Pasteur's Quadrant: Basic Science and Technological Innovation* (Washington, DC: The Brookings Institution).

Verloo, Mieke and Sylvia Walby (eds) (2012) 'Intersectionality in the Equality Architecture', Special Issue *Social Politics* 19 (4) : 433–621.

Weldon, S. Laurel (2002a) 'Beyond Bodies: Institutional Sources of Representation for Women in Democratic Policymaking', *Journal of Politics*, 64 (4), 1153–74.

Weldon, S. Laurel (2002b) *Protest, Policy and the Problem of Violence Against Women: A Cross National Comparison* (University of Pittsburgh Press).

Weldon, S. Laurel (2008) 'Intersectionality', in Goertz and Mazur (eds) *Politics, Gender and Concepts: Theory and Methodology* (Cambridge University Press), 193–218.

Weldon, S. Laurel (2011) *When Protest Makes Policy* (Ann Arbor: The University of Michigan Press).

Zippel, Kathrin (2009) 'The European Union 2002 Directive on Sexual Harassment: A Feminist Success?', *Comparative European Politic*s, 7 (1), 139–57.

11
Mixed-Methods Designs in Comparative Public Policy Research: The Dismantling of Pension Policies

Sophie Biesenbender and Adrienne Héritier

Introduction: The methods of mixing research approaches

Mixed-methods designs have received burgeoning attention in the academic community during the last decade not only in the social sciences but also in public health research and psychological science (Giddings, 2006; Doyle et al., 2009). Multimethod approaches and techniques of triangulation have a long tradition in these literatures (Campbell and Fiske, 1959; Jick, 1979; Brewer and Hunter, 1989). They have however only recently been translated into a 'unique research approach that has philosophical foundations, its own terminology, systematic research designs, and specific procedures for designing, implementing and reporting research using this approach' (Plano Clark et al., 2008: p. 354; see Greene, 2008). Apart from the feature to combine qualitative and quantitative approaches, mixed-methods research differs from multimethod designs by the interwovenness of the underlying research questions that can only be answered through different analyses (Johnson et al., 2007; Morse, 2010).

From the single-method perspective, the choice of either a qualitative or a quantitative method of approach implies gains and losses. There is a trade-off between the gain of more in-depth insights and the loss of an overall view of the phenomena of interest in the first case, and vice versa in the second case (see Coppedge, 1999; Fearon and Laitin, 2008; Caporaso, 2009). Hence, the decision to choose one approach instead of the other largely depends on the research question posed and whether it may be better grasped by using qualitative or quantitative data (see Creswell, 2003).

Choosing a qualitative approach will allow gaining in-depth insights into relatively complex political phenomena, but will impose limits on

what one can generalize from the empirical cases studied. Choosing a quantitative approach, inversely, will allow for generalization across a large number of phenomena, but will not allow insights into the specifics of a phenomenon and any underlying causal processes between *explanans* and *explanandum*.

However, the choice between the use of quantitative or qualitative data need not be an absolute one. Rather, both approaches can be linked in a mutually beneficial way (Lieberman, 2005). Thus, starting from a large-N analysis allows a limited number of case analyses to control whether all important factors have been included. In the reverse sense, single case analysis offers large-N analysis the possibility to control for spurious correlations and endogeneity by analyzing the underlying causal mechanisms.

In this context, the mixed-methods strategy offers ways to mitigate the weaknesses of the two broad research approaches within the social sciences (see Coppedge, 1999; Gerring, 2005; Mahoney and Goertz, 2006; Denscombe, 2008). While quantitative analyses are apt to ensure the generalizability of the results, additional qualitative case studies of the processes might help to identify the causal mechanism at work for a subset of the units of analysis.[1]

The value addition of mixed-methods over single-method approaches lies in the ability to draw more robust conclusions from empirical research. At the general level, one can conceive of four different mixed-methods designs; each of which enhances the credibility of the research findings (Table 11.1; see Teddlie and Tashakkori, 2006 for alternative classifications of mixed-methods approaches). While two of the four constellations eventually lead to more substantiated results (that is, the purely exploratory and confirmatory approaches), the remaining two set-ups are appropriate when extensively studying, for instance, under-researched topics (cf. further Morse, 1991, 2010; Hall and Howard, 2008; Creswell, 2010).

On the one hand, *complementary* uses of different methods (either confirmatory or exploratory) are most appropriate to enhance the validity of findings and therefore to minimize the risks of biased results (Coppedge, 1999). On the other hand, the *sequential* application of different methods is apt to deliver both the development of a theory and its subsequent testing. This strategy helps to obtain generalizable results by establishing close empirical links in both theory-building and testing.[2] Ultimately, the decision of whether to choose a mixed-methods design at all and whether to focus on complementary or sequential designs depends on the researcher's individual philosophical point of departure.[3]

Table 11.1   Typology of mixed-methods approaches

		Quantitative perspective	
		Exploratory	**Confirmatory**
Qualitative perspective	**Exploratory**	*Complementary* • Extensive theory-building (hypothesis generation)	*Sequential* • Qualitative identification of causal mechanisms • Subsequent quantitative testing of hypotheses
	Confirmatory	*Sequential* • Quantitative identification of causal mechanisms • Subsequent qualitative testing of hypotheses	*Complementary* • Extensive theory-testing (hypothesis testing)

While complementary mixed-methods approaches allow for comparing results obtained through the application of different research techniques, emphasis in sequential methods is put on the validation and explanation of the findings obtained by distinct approaches. First, qualitative in-depth studies subsequent to a large-N analysis might validate both the constructs underlying the different research methods (Hitchcock and, Nastasi, 2011) and the data obtained through distinct analyses (Plano Clark et al., 2008). In the context of survey research for instance, this strategy allows researchers to scrutinize quantifiable indicators by adding qualitative open-end questions. Second, for explanatory purposes scientists can choose to complement large-N analyses with in-depth qualitative case studies in order to identify the precise mechanisms at work – such as through in-depth process-tracing (see Caporaso on analytical as opposed to descriptive narratives, 2009) or grounded-theory approaches (cf., for example, Corbin and Strauss, 1998; Creswell, 2003). Similarly, they might also perceive the need to understand why there are deviations from certain empirical or theoretically predicted patterns (see Gerring, 2007, for different ex-ante and ex-post case study techniques) in certain cases. In this situation, follow-up qualitative case studies might enhance the findings obtained from large-N analyses. Third, the reasons for adding qualitative analyses to a quantitative one might be given by the researchers' objective to identify the underlying micro-macro link (Kittel, 2006) in order to enhance the credibility of quantitative research results. Fourth, researchers might seek to confirm qualitatively

obtained results based on a broader sample for reasons of generalization (see King et al., 1994; Coppedge, 1999; Mahoney and Goertz, 2006; Gerring, 2007; Mahoney, 2008). A fifth and more research pragmatic reason for mixing different methods is the need to find a selection criterion for qualitative in-depth analyses. In this context, ex-ante quantitative studies are suitable to detect empirical scenarios and types of interesting cases. In the literature a number of techniques have emerged to select cases for follow-up qualitative analyses (see Seawright and Gerring, 2008).

In short, mixed-methods studies can be motivated by both research-pragmatic and substantive theoretical reasons. Further, their results are more persuasive as potentially they reach a broader group of scholars who pertain to different research schools and traditions (see Plano Clark et al., 2008). Given the broadened perspective associated with mixed-methods designs when compared to single-method approaches, it is hard to conceive of any conceptual or analytical reasons against these kinds of research designs. Rather, the actual challenge consists in the combination of different methods in research practice both by exploiting the strengths and by mitigating the weaknesses of different approaches.

This chapter discusses individual research steps that are taken when conducting comparative public policy research. It starts from the most elementary query of what questions that constitute a puzzle in comparative public policy research are worthwhile studying and asks how this puzzle is transformed into a systematic research question. Two different ways may be chosen to look for answers to the research question: deducing hypotheses from existing theories in the field of investigation; or exploring empirical material and research – in an inductive way – for possible new answers to the research question. Once hypotheses have been formulated, their independent and dependent variables and the underlying causal mechanism are specified. In a further step, the independent and dependent variables are operationalized; that is, empirical indicators are defined that may be used to measure the presence or absence of the variables in our empirical material. Finally, we have to describe the quantitative and qualitative data and the sources that we are using to collect them, how we systematized and evaluated the data and what our empirical findings tell us regarding the validity of our hypotheses. To give a practical example, we exemplify all of these steps by describing a concrete research project focusing on the dismantling of pension policies in Western democracies, and more particularly in Italy and Switzerland, and discuss the respective advantage of choosing a quantitative or qualitative approach when collecting data and evaluating them.

Empirical puzzle and research question: Why adopt a mixed-methods research design?

Before engaging in a research project on comparative public policy, the researcher clearly states the added value of planned research, anticipating the famous 'so what' question. The added value may lie in an added empirical *or* theoretical value of the planned research. Has the researcher been motivated by some fascinating variation in empirical political outcomes, a striking empirical inconsistency, or a surprising new empirical development? Or does it start from a theoretical puzzle, an on-going theoretical controversy or inconsistent arguments in the literature? A mixed-methods approach allows one to define a research question from two different angles: the qualitative research is based on acquiring in-depth knowledge of the empirical phenomena in question. This in-depth knowledge may lead to the discovery of new views on the topic in question not covered by the existing literature. The qualitative view is not restricted by the availability of existing quantitative data or by the need to be able to quantitatively measure the phenomenon in question and the concomitant need to narrow down the perspective to what is quantitatively measurable. It has to generate new data by using the corresponding empirical tools. The advantage of a quantitative approach consists in pointing to common limited features across a large number of cases, and, therefore, may offer useful insights as to where individual qualitatively researched cases are situated with respect to this feature compared to other cases. It is for these reasons that we chose two distinct analyses, qualitative case comparisons and a quantitative analysis across a large number of countries to account for the conditions and modes of dismantling pension policies.

Once the puzzle that motivates the research has been described, the research question is formulated in a systematic way. In qualitative research it usually takes the form of a 'why' question: Why has a particular phenomenon emerged, or changed over time, and what are the consequences of this change? In other words, we pose causal questions.

In our empirical example we focus on the puzzle of how policies are dismantled and why the mode in which they are dismantled varies across countries. While the large-N perspective allows assessing and predicting the overall extent of policy dismantling over a number of cases, the qualitative case-study based strategy is useful to study the particular dismantling strategies pursued by the policy-makers. Put differently, the qualitative search for causal mechanisms is complemented by the quantitative assessment of the average changes in the dependent variable for the units of analysis over a broad geographical and temporal sample.

Pension policies constitute an important cost factor for public budgets and are opposed by important and powerful constituencies in demographically aging societies. In view of these features how do various states deal with the problem of reducing entitlements to pensions? While considerable research has been conducted on growth and innovation in social policy, the study of the conditions and modes of abolishing existing policies was – for a long time – little investigated (Bauer and Knill, 2012). We therefore formulate our research question as follows: 'To what extent do states reduce pension payments, and under what conditions and by which political processes?' This is not only empirically important, but also theoretically challenging. Whereas the overall pressure from demographic and economic developments is relatively straightforward, the question of why states react in different ways politically to this pressure and opt for specific modes of dismantling is theoretically more challenging (Bauer and Knill, 2012).

In the light of the above-discussed types of different mixed-methods designs and existing literature on policy change in the field of pension politics, we opt for a complementary theory-driven research design. Put differently, we rely on both qualitative and quantitative research techniques to evaluate theoretically derived hypotheses on policy dismantling in a distinct policy field. These refer to both, the average effects of certain possible factors on the governments' willingness to reduce generosity and the precise mechanisms at work that account for these decisions. Integrating qualitative and quantitative research methods we are able to analyze both the causes and conditions of policy dismantling in the field of pension politics.

Concepts and theories: How to elaborate a mixed-methods research design

Generating hypotheses

Causal questions call for answers or hypotheses which claim that specific factor(s) (independent variable/s) explain a specific outcome (dependent variable). They may be generated in two different ways. They may be derived from existing theories and as such offer an answer to the research question of interest (deductive procedure). Or hypotheses may be gained inductively from prior empirical exploratory research which produced interesting new causal insights between classes of phenomena. When proceeding deductively, the theoretical reasoning starts out with an explication of the assumptions on which the theoretical argument is based. The assumptions indicate what is assumed as given, such

as actors' motives of behavior or a multilevel structure of governing. In this context, sequential mixed-methods designs constitute powerful means to combine inductive exploratory approaches with subsequent deductive theory-driven empirical testing (see *Introduction*).

On the basis of the explicated assumptions, a systematic variation of either the explanatory macro variables or the micro variables emphasized in the theory produces hypotheses that state a causal claim. The researcher opts for a theory, which in her view offers the most explanatory power. She may also propose two competing theoretical explanations which predict different outcomes; or different theoretically derived explanations may be offered which are complementary in that one theory explains one part of the variance of the dependent variable, and, the other theory, the rest of the variance. Finally, two theoretical explanations may also be observationally equivalent or equifinal in that they predict the same result but claim that the same outcome may be reached through two different processes.

Whether proceeding in a deductive or inductive way, a causal statement should also explicate the causal mechanisms – preferably expressed in theoretical terms – linking the *explanans* to the *explanandum*. Explicating causal mechanisms helps discover spurious relationships in the correlation analysis of covariation, which has been critiqued for sometimes producing significant correlations between nonsensical phenomena. In that sense, correlations do not really explain an outcome. However, the contrast should not be overemphasized. For one thing, a causal explanation based on qualitative data and case studies does not make sense without the link between an independent and dependent variable. Second, correlation analysis may consider the underlying causal mechanism as so obvious that it does not warrant explication (Gerring, 2005).

Mixed-methods approaches go beyond single-method research in that they allow for the integrated analysis of the causes and conditions for certain empirical phenomena. Our analysis of policy dismantling follows a complementary deductive research design that seeks to combine qualitative case studies with quantitative regression analysis in a complementary way (cf. further Onwuegbuzie and Leech, 2006). In addition to enhancing the validity of the data and the research results (see *Introduction*), this strategy allows combining different logics of linking the independent and dependent concepts (that is, deterministic versus probabilistic approaches; see Neuman, 2002; Brady and Collier, 2004; Tarrow, 2004; Mahoney, 2008). While the quantitative analysis follows the effects-of-causes approach, qualitative studies are deterministic in

nature in that they seek to identify necessary and sufficient causes for certain empirical phenomena (see Holland, 1986; Ragin, 1989; King et al., 1994; Sobel, 2000; Mahoney and Goertz, 2006). With regard to our concrete example of pension policy dismantling, the integrated mixed-methods analysis enables us to distinguish between conditions that facilitate and those that actually cause the dismantling processes.

In the qualitative analysis of our concrete example of why pension policies develop in different ways across different countries, we generate hypotheses from rational choice institutionalism and bargaining theory. Based on strategic interaction analysis (see Frieden, 1999; Lake and Powell, 1999) we distinguish between a micro or actor's side on the one hand and a macro or environment/institutional side of the analysis on the other. We systematically vary either micro aspects while controlling for macro features in order to predict outcomes, or vary macro features while controlling for micro factors on the other, in order to predict the outcomes of a strategic bargaining process among the relevant actors in a specific macro context.

On the micro side we assume that for relevant political actors, cutting back pension payments is politically costly since they may subsequently be 'punished' by a loss of votes. However, while all actors shun these costs, they are also under pressure to decrease budget deficits, reduce social contributions and make sure that the younger generations are not too heavily burdened with pension payments. Hence, we assume that the relevant political decision-makers, while all of them are under pressure to cut pension costs and agree that measures need to be taken, have diverse preferences as regards how to proceed to reduce payments. We assume that they have a variety of possible strategies at hand, but have different preferences over the ranking of these strategies. These outcomes or strategies are: (i) a manifest considerable cutback; (ii) a hidden or automatic incremental cutback by indexing or non-adjustment to rising living costs; (iii) a shifting of the problem solution to another level; and (iv) no action at all.

On the macro side we distinguish environmental or macro factors which we assume to be relevant for the outcome of actors' strategic bargaining. It is plausible to assume that the institutional structure of a political system, for example, the number of veto players in a political system and the number of levels across which decision making extends, are the most important political macro variables.

In a first conceptual experiment[4] we assume high problem pressure and diverse actors' preferences as regards outcomes. We vary environmental/macro conditions to predict a corresponding outcome of the

strategic bargaining process: '*The more veto players with diverse preferences in the political arena, the more incremental (or hidden) dismantling strategies will be.*' (H1)

We argue that given high problem pressure, diverse actors have different preferences as regards the strategies of cutting back pensions schemes. Whilst left-wing parties are expected to oppose cutbacks in public pension systems, center-right parties, in order to safeguard the economic viability of the public pension system, may be more willing to engage in cutbacks urging citizens to rely on private pension schemes. Both parties will be under considerable pressure from the respective interested organizations representing their constituencies to resist or support cutbacks. In an institutional set-up formal and de facto veto players have considerable clout in bringing their influence to bear. To put it differently, where a veto-player can prevent action from being taken, it is plausible to expect that dismantling will be difficult to carry out and can only be pursued in an incremental and less noticeable way.

In a second conceptual experiment we vary actors' preferences while controlling for institutional macro conditions and hypothesize: '*Under the condition of a strong economic downturn and financial pressure of public budgets, ceteris paribus, political actors will opt for straightforward cutbacks in pension payments.*' (H2)

In a situation where problem pressure increases due to a stark economic downturn and increasing financial problems in the public budget – and assuming the same institutional macro conditions in both cases – we may expect that the urgency of the situation will influence actors' preferences over the desirable outcomes. Situational preferences (if not the meta preferences) of actors may change and induce more willingness to pursue straightforward cutbacks in pension payments.

Quantitative regression analyses are useful in assessing the mean causal effect of one factor on another by approximating the measurement of causes with that of correlations. As such, most quantitative studies share the assumption that causes are additive in nature with the exception of those approaches that explicitly model the expected interaction effects of the different factors. Therefore, the quantitative approach requires hypotheses that express the expected average change in the dependent variable changes as a reaction to modified independent factors. In the context of confirmatory analyses, the theoretically identified causes and mechanisms translate into a variety of different testable claims. On the one hand, hypotheses might express the expected direction of change in the dependent variable as a response to certain developments of the independent variables. On the other hand, they

can state the relative impact of the different independent factors in order to compare the coefficients' values and levels of significance (given comparable metrics of the underlying indicators).

Quantitative analyses include explanatory factors of various natures. On the actor-centred side, we expect the actors' preferences to be of relevance for political decisions. While large-N analyses are an adequate means to assess a government's general willingness to engage in policy dismantling, they hardly allow for the distinction between the different strategies of policy dismantling (see above Hypothesis 1). From this perspective, actors' preferences with regard to pension policies are a function of the political ranking of the issue. Therefore, the actors' decisions to reduce generosity of the pension scheme positively correlate with their expressed willingness to implement cutbacks (with respect to, for example, old-age insurance).

We hypothesize: '*The stronger the actors' (legislative branch) average prioritization of cutbacks (in the generosity of the social security scheme), the higher the actors' propensity to engage in policy dismantling*' (H3)

On the environmental/institutional side, we may consider variables to measure the degree of problem pressure the decision-makers are exposed to. Variables for this theoretical concept include the financial pressure for the pension scheme in general. We assume that decisions to reduce generosity of the pension scheme are positively related to the expenditure of the scheme. Based on this distinction, we may derive the following hypotheses to test within a large-N analysis of policy dismantling in the field of pension politics: '*The higher the financial pressure of the pension scheme, the higher the actors' propensity to engage in policy dismantling*' (H4).

Specifying hypotheses and operationalizing concepts

The hypotheses under which both the qualitative and quantitative approaches are initially stated in general terms, need to be specified with respect to the specific content of their independent and dependent variables. Empirical indicators are used as measures of the phenomenon to be empirically assessed, both on the side of the independent as well as dependent variables. An empirical indicator of a phenomenon of a political reality constitutes one possible measure of the phenomenon in question and does not claim to represent the latter in an exhaustive way. Frequently, in order to better grasp an empirical phenomenon, several empirical indicators are chosen and aggregated in an index.

In order to specify the hypotheses in the qualitative analysis, '*The more veto players with diverse preferences in the political arena, the more incremental*

(or hidden) dismantling strategies will be.' (H1) and *'Under the condition of a strong economic downturn, ceteris paribus, political actors will opt for straightforward cutbacks in pension payments.'* (H2), we have to specify how we measure veto players with diverse preferences (V ind) and incremental and hidden dismantling strategy (V dep). Veto players are defined as formal and informal veto players, that is, actors who, owing to their formal competencies in the political decision-making process and/or due to their economic, social and political power, may block the formal adoption of a policy change. Incremental and hidden dismantling strategy is defined as a strategy which implies a small cutback of pension provisions and, for example, is linked to an indexation or other form of automatic adjustment, instead of a straightforward legislative reduction of provisions (H1). Strong economic downturn is defined as zero growth of GDP or shrinking of GDP and concomitant financial problem pressure (V ind). Straightforward cutbacks in pension payments are defined as a legislative amendment providing for a cutback in pension payments and measured in *de facto* cutbacks over time.

In the quantitative analysis, in order to allow for empirical confirmation or disconfirmation of the hypotheses – *'The stronger the actors' (legislative branch) average prioritization of cutbacks (in the generosity of the social security scheme), the higher the actors' propensity to engage in policy dismantling' (H3)* and *'The higher the financial pressure of the pension scheme, the higher the actors' propensity to engage in policy dismantling' (H4)* – the theoretical concepts need to be translated into measurable indicators. Indicators are the link between the theoretical concepts and the empirical measure. The choice of indicators is a fundamental decision with immediate impact on eventual regression results. Therefore, this research stage requires particular attention in order to ensure the high-quality measurement of the underlying theoretical concepts.

The dependent variable in our empirical example is convincingly operationalized by counting the negative changes in the policy output over time (Knill et al., 2010; see above). More specifically, policy dismantling covers the number of instances in which political decision makers engage in generosity-reducing measures over countries and time. Knill et al. (2010) propose a classification scheme of policies in order to assess the different dimensions of policy dismantling. According to the analytical framework, activities can, on the one hand, affect the density of the policy field by abolishing entire policies or instruments. On the other hand, it is possible that policy dismantling targets the intensity of policies by either lowering the level of generosity (for example, the reduction of pension payments) or by tightening the scope of regulation

(through more stringent eligibility criteria, for instance). Therefore, the dependent variable to assess the concept of policy dismantling in public pension schemes includes the number of policy dismantling instances per year over countries and time. In a more refined analysis, these measures may be classified according to the two above-mentioned broader categories (density and intensity) or the four related subtypes (policy, instrument, level, scope). Therefore, the count data for these four categories of dismantling sophisticatedly mirror the dynamics in the pension policy field in 24 OECD countries over time (on a yearly basis spanning 30 years).

The operationalization of the independent variables is less straightforward. On the micro side, the actors' average prioritization of cutbacks can be approximated by taking the parties' average inclination towards the reduction of the welfare state and public generosity weighted by their respective stake in the decision-making process (that is, their degree of representation in parliament). The Comparative Manifestos Project (CMP) dataset (Budge et al., 2001; Klingemann et al., 2006; Volkens et al., 2011) provides measures on the parties' willingness to reduce social benefits as well as on their role in the legislative branches of governments. The sum of the weighted party positions, according to the relative number of seats in parliament, yields an aggregate measure of the parliament's endorsement of cutbacks in public spending for social policies. On the macro side, financial problem pressure can be captured by the development in government expenditure for the public pension scheme over time and countries. The indicators for this variable comprise measures of the total outlay of public pension schemes over time including expenditures for cash benefits and benefits in kind. A rather indirect measure of a country's financial leeway for keeping up its generosity in the public pension scheme is approximated by the overall growth of the economy (GDP). This general measure of the economy's trend is indicative of a country's ability to balance adverse demographic developments. We expect a positive relationship between the degree of policy dismantling of the public pension scheme and the parliament's average inclination towards welfare state retrenchment. Government expenditure for the public pension scheme is also predicted to have a positive effect on the dependent variable, while general economic growth (increase in GDP) is most likely to have a negative impact.

In addition, we included a number of control variables that refer to the degree of problem pressure (the share of persons aged 65 years and more), electoral opportunity structures (that is, the timing within the electoral cycle) as well as the regulatory style of the current legislature

(that is, the rolling average of dismantling events observed so far per year in the on-going legislature).

Altogether, the research design of our empirical example is centred toward the integrated analysis of two micro hypotheses (H1 and H3) and two macro hypotheses (H2 and H4) based on the integration of case studies and regression analyses. This strategy allows compensating the deficits of one method with the strengths of another and helps to increase the validity of the empirical data. More specifically, we would gain more robust insights into the causes and conditions of policy dismantling in a specific policy field.

Measurement: Data collection in mixed-methods research

Empirical indicators may be expressed in quantitative or qualitative terms. Whether to use a large-N quantitative analysis or study a limited number of cases based on qualitative data depends on the research question we seek to answer.

Qualitative data

There is a variety of tools to be used when collecting qualitative data, such as conducting interviews, analyzing documents, or engaging in participatory observation. The decision over which methods to use is guided by the question: Which data do I need in order to measure the value of the independent and dependent variables of my hypotheses? Which data do I need in order to measure the causal process leading from my independent to my dependent variable? Depending on what these values are information will be collected from existing documents, or through interviews or participatory observation. If possible several instruments should be combined in order to crosscheck the data which have been collected.

When confronted with a collection of qualitative data, whether collected in interviews, which were taped and transcribed, or whether taken from official documents, researchers face the daunting problem of how to process such piles of information (Miles and Huberman, 1994; Corbin and Strauss, 1998). Two basic approaches may be chosen to initiate this process: a deductive approach in which the data analysis will be guided by the conceptual framework, the hypotheses, key variables and empirical indicators, which were formulated to answer the research question; or by using an inductive approach in which the researcher uses a general accounting scheme, with no prior precoding, using only elementary notions such as the overall setting and the definition of the

situation and strategies used, thereby allowing qualitative codes to 'emerge' from the field. The objective is to generate entirely new hypotheses suggested by the empirical findings. Here we will focus on the more frequently used hypothesis-guided, deductive approach.

The empirical indicators which were formulated when operationalizing the variables of the research hypotheses are used to identify the presence and absence of data reflecting these indicators in the data material. The empirical indicators are the codes used to sort the qualitative data. Codes are tags or labels for assigning units of meaning, for example, words, sentences and paragraphs, to the data. The codification of the empirical material is a condensation of data from the viewpoint of the pre-defined empirical indicators in order to identify information confirming *and disconfirming* the expected values on the variables (Miles and Huberman, 1994; Corbin and Strauss, 1998).

Coding allows the researcher to organize chunks of data so that segments relating to the research hypotheses can be found, pulled out and clustered. When the investigator goes through the transcripts or documents, units dealing with the same topic/empirical indicator and subtopics that recur with some regularity are marked off (Miles and Huberman, 1994; Corbin and Strauss, 1998). It is crucial that the empirical material is searched and classified until all empirical indicators – in the sense of a confirmation and disconfirmation – of the research hypotheses have been identified and used for validation of the latter. The coding or evaluation of the empirical material in the light of the empirical indicators of (dis)confirmation, should be conducted by two or more researchers working independently of one another.

Quantitative data

Quantitative research needs to rely on less detailed and often times more remote indicators for the operationalization of the theoretical concepts. This lack of specificity is due to the need to standardize the empirical measures in order to apply them to a sufficiently broad sample. This requirement makes it even more difficult to identify appropriate indicators that balance the claim of universal applicability on the one hand and the necessity to accurately reflect the underlying theoretical concept on the other.

The process of data collection is a complex, personnel-intensive and time-consuming endeavour that, in most cases, requires the collaboration of a number of researchers. Therefore, single researchers often rely on available data records, or in a combination thereof, in order to be able to individually conduct large-scale quantitative data analyses. In other cases, independently collected data complement existing databases.

Regardless of the specific origin, the reliance on external data sources always bears the risk of inaccuracy or low levels of data quality. This might lead to systematically biased estimates in the event that the deviances are not at random. In certain cases, quantitative estimation techniques are apt to handle some of the problems associated with poor data quality (for example, missing values). The handling of external data is further complicated by the limited information as regards the detailed process of data collection and the problems associated with the coding of mostly qualitative information into quantitative measures. Moreover, external data often reduce the researcher's flexibility with regard to the operationalization of the theoretical concepts.

Empirical data analysis: Pension dismantling in Italy and Switzerland

The goal of our study is to explain the occurrence of policy dismantling in pension politics by means of an integrated research design that addresses two sets of hypotheses (micro and macro) with the help of both qualitative and quantitative research techniques. The following paragraphs describe the empirical results that have been obtained in the course of the complementary case study and regression analyses. It concludes with an integrated discussion of the results and the evaluation of the analysis-specific hypotheses in light of findings obtained by the alternative method.

In order to gain empirical insights from the qualitative case study, we compare two attempts to reform the Italian and Swiss public schemes respectively. These precise cases are of particular theoretical and empirical interest as the countries have experienced constant reform pressures and initiatives over the last two to three decades. Despite considerable institutional differences between Switzerland and Italy (which implies distinct veto-player constellations), policy-makers in both countries have had to face strong politicization and social fragmentation over the issue of pension reform. In the past, this has also led to failed reform initiatives – either due to the public vote (in the Swiss case) or social mobilization (in the Italian case).

In the qualitative analysis in order to empirically assess H1, '*The more veto players with diverse preferences in the political arena, the more incremental (or hidden) dismantling strategies will be*.' (H1), we compare two cases in Italy: the Berlusconi reform of 1994 and the Dini reform of 1995 (Schmitt, 2012). Under pronounced financial pressure on the public pension system, the Berlusconi government sought to realize pension cuts to

dismantle the seniority pensions (allowing workers to retire after 35 years of work). The government – backed by the employers' organization – sought to implement drastic pension cuts and the abolishment of seniority pensions. The trade unions as *de facto* veto players mobilized strong political opposition and organized nation-wide strikes, which gradually threatened the internal cohesion of the government's parliamentary support. The government signed an agreement with the unions renouncing most of the cutback measures (including the seniority system). Shortly afterwards, the Berlusconi government resigned. In other words, a straightforward legislative cutback of pension provisions had failed (Schmitt, 2012). Under the subsequent technical government of Dini, all potential *de facto* veto players were brought to the table to negotiate pension reform – although the employers' association left the bargaining table at one point. The outcome was a watered down, incremental reform of gradually phasing out entitlements, such as seniority pensions. A range of local referenda were held which produced majority support for the reform proposal (Schmitt, 2012). The comparison of the Berlusconi 1994 reform and the Dini 1995 reform suggests a confirmation of H1 that if diverse veto players with diverse preferences negotiate pension cutbacks, the outcome will be moderate and incremental.

The Swiss reforms based on consensual decision-making processes among all political parties and social forces provide additional empirical information that net direct dismantling – even under fiscal pressure – does not take place. The 1994 reform law, for instance, shows how incremental realized dismantling is and how it was embedded in comprehensive packages that included elements of additional provisions of pension rights (to women), rendering cutback measures less obvious. While the reform law mitigated the negative discrimination of women (by introducing a contribution splitting system for spouses, for example), it also abolished elements of positive discrimination by increasing the age of retirement for women from 62 to 64 years (Bonoli, 2007). While in the Swiss case, the 1994 reform law entailed dismantling elements under the disguise of introducing gender equality to the public pension scheme, the 2003 reform law was explicitly designed to adjust the insurance scheme to meet future financial and demographic challenges. Nevertheless, the eventual policy package of 2003 was carefully balanced between elements of policy dismantling and compensatory expansion. Interestingly, the measures of dismantling were more remote in character, while the expansion elements had immediate consequences for the retirees. A range of expenditure-decreasing policies (such as gradual increases in contribution rates, cutbacks in pension

indexation, gradual increase in value added tax rates to provide financial backing to the scheme and elimination of privileges for employed retirees) were of quite remote character. There was only one particularly evident element of dismantling that regarded the full harmonization of female and male ages of retirement; hence an increase of one year for women was, however, countered by generous and subsidized early retirement provisions that actually constituted an incentive to retire before reaching the official age. Therefore, the incremental dismantling processes of the Swiss 1994 and 2003 reform laws constitute a further confirmation for hypothesis H1.

In order to empirically probe H2, '*Under the condition of a strong economic downturn, ceteris paribus, political actors will opt for straightforward cutbacks in pension payments*', we compare the Italian and Swiss cases in their entirety. More precisely, the findings suggest that the degree of financial problem pressure does not have an immediate effect on the content and direction of the eventual policy package itself. While constituting a necessary condition for the dismantling process to set in, policymakers seem to be guided by other motives when selecting the strategies of dismantling. In the Italian case, membership in the Economic and Monetary Union and the risk of failing to meet the EU convergence criteria by 1998 constituted the eventual (external) incentive for the governments in the first half of the 1990s to reduce public spending and to economize the welfare protection schemes (Ferrera and Jessoula, 2007). Even though this situation triggered a series of attempts to reform the public pension schemes in Italy (starting in 1992), governments have continuously opted for incremental and hidden cutbacks in generosity (with the exception of the failed Berlusconi reform).

The Swiss cases show that internal political developments explain the timing of reform laws. The idea of restructuring the pension scheme regained attention due to changes in the political administration. Flavio Cotti, who became secretary of the interior in 1987, acted as a political entrepreneur furthering not only the gender-equality issue that led to the 1994 law but also initiating the financial adjustment process that accounts for the expenditure-decreasing policies of the 2003 reform law (Bonoli, 2007). In spite of the generally perceived need for reform, this initiative did not lead the veto players to agree on ambitious dismantling policies, which was also due to the risk of defeat by referendum. As a consequence, the Swiss dismantling policies have also been of incremental and hidden character in the past.

The Italian and Swiss cases show that a lack of economic growth and financial pressure on the public pension systems alone do not lead to

up-front cutbacks in the pension system. Rather, financial pressure seems to be a necessary condition which, when linked to other factors – such as the number of *de facto* veto players – leads to cutback strategies. Hence, H2 is disconfirmed.

The quantitative data on policy dismantling in the public pension scheme form part of the European Commission-supported CONSENSUS project (see Knill et al., 2010). The development of the dependent variable aggregated over 24 OECD countries is plotted in Figure 11.1. In this context, the abolishment of policy instruments or entire policies is summed up under the label of *policy density*. Reductions in levels and scope of pension policies are reflected in the *policy intensity* indicator. The comparison between the two indicators illustrates that countries tend to scale down expenditure in the public pension scheme by reducing the generosity of certain policies rather than by restructuring the whole pension scheme. The entire abolishment of existing policies or instruments constitutes an exception. More precisely, policy-makers prefer to cut back the amounts of old-age pensions or to tighten eligibility criteria (for example, by increasing the age of retirement) instead of reducing the number of parallel measures. This finding is intuitively
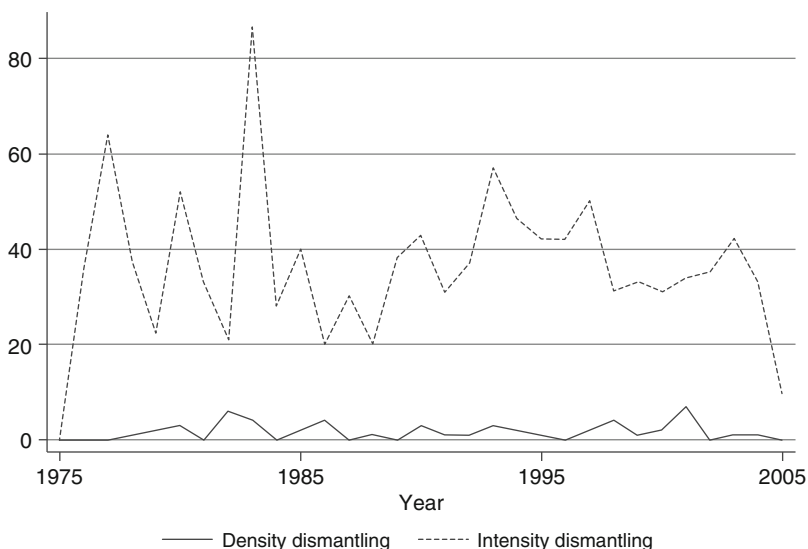

Figure 11.1   Dismantling in the public pension scheme: Aggregated count data for 24 OECD countries
Source: Knill et al., 2010.

understandable if we assume the underlying blame-avoidance strategies of the policy-makers. Given that dismantling of policy density (instruments and policies) implies the simultaneous dismantling of policy intensity (levels and scope), the former type of activity constitutes the more extensive and also more visible alternative of generosity reduction. With regard to the calibration of pension policies (that is, the dismantling of levels or scope) the data show strong fluctuations in the whole sample of OECD countries over time with strong peaks, especially during the decade from 1975 to 1985. In spite of the absolute numbers of dismantling having moved within a narrower margin since 1985, the aggregated data do not show a clear upward or downward trend.

The empirical picture for Switzerland and Italy roughly confirms the evidence drawn from the case studies. Figure 11.2 plots the country-specific numbers of policy dismantling instances of any type over time. Both countries exhibit moderate fluctuation of dismantling activities over the period under consideration (1975 to 2005) except for the strong peaks of policy dismantling that affected the two countries during the first half of the 1990s (see Figure 11.2). The descriptive data illustrate that pressure for dismantling the public pension scheme in Italy and Switzerland was strongest during the periods covered by the qualitative analysis.
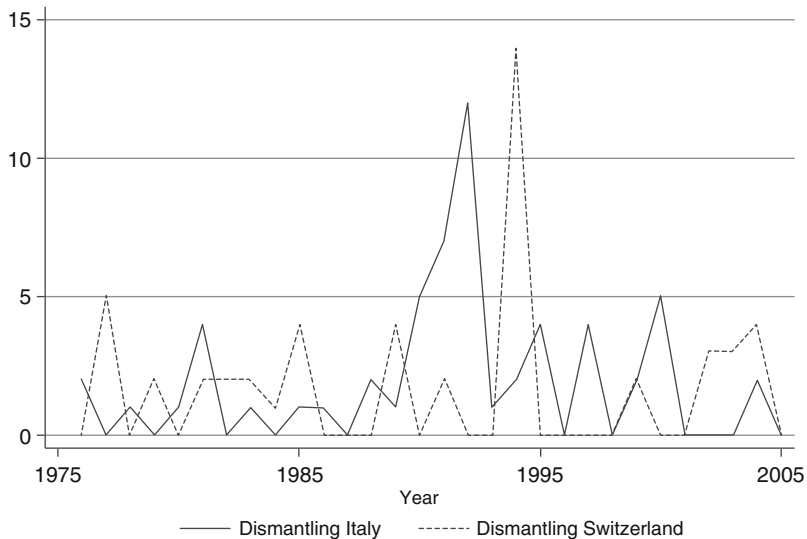

Figure 11.2   Dismantling in the public pension scheme: Count data for Switzerland and Italy
Source: Knill et al., 2010.

This descriptive evidence on policy dismantling cannot be exhaustively analyzed without reliance on external data. Quantitative data analysis usually requires the use of additional data in order to provide satisfactory grounds for statistical inference. In this context, a number of external data sources have been used to gather information on the independent and control variables for the subsequent analysis. The actors' prioritization of welfare state retrenchment (Hypothesis 3; see above) can be assessed by relying on the relevant indicator of the CMP dataset (Budge et al., 2001; Klingemann et al., 2006; Volkens et al., 2011). This indicator yields the percentage of (quasi-) sentences expressing the endorsement of *welfare state limitation* in the parties' electoral manifestos. This measure constitutes the basis for estimating the development of parliament's prioritization of cutbacks in public generosity (see *Measurement* section). The indicators for the independent variables from hypotheses H4 can be gathered with the help of OECD statistics that are available from 1980 onwards. The databases provide estimates for the development of government expenditure for the public pension scheme in OECD countries over time. Moreover, they offer a wide array of GDP estimates based on different methods and estimation techniques. Among the control variables, we included measures for problem pressure (World Bank, 2012), the electoral cycle (based on the *Democracy Time-series Dataset;* Norris, 2009) and regulatory style (which refers to the rolling annual mean of dismantling decisions in the ongoing legislature based on CONSENSUS data and the *Comparative Political Data Set I;* Armingeon et al., 2011).

It is beyond the scope of this chapter to deliver complete and encompassing testing of a general theory of policy dismantling. Rather, the following regression results serve to explicate ways in which an empirical phenomenon can be approached from different methodological perspectives by discussing and testing selected hypotheses. Therefore, emphasis is placed instead on comparing the complementary evidence obtained with the help of two distinct techniques of analysis (qualitative and quantitative).

Table 11.2 provides primary empirical evidence for the above-deduced hypotheses. It illustrates the correlations between different operationalizations of the dependent variable (policy dismantling of public pensions) and the theoretically identified explanatory factors by means of a zero-inflated Poisson regression (with the variable for regulatory style approximating the degree of zero inflation). The estimates are reported in the incidence rate ratio format, which eases the interpretation of the coefficients. More precisely, the rate ratio of the dependent variable changes by the estimated factors if the independent variables are to

Table 11.2 Dismantling of pension benefits in OECD countries, 1980–2005 (zero-inflated Poisson regression)

		Overall dismantling	Policy intensity Dismantling		
		Total	Total	Level	Scope
Main					
Political parties	Support for welfare dismantling by political parties in parliament (CMP values weighted by number of seats in parliament)	1.0837* (0.051)	1.1279** (0.055)	1.2360*** (0.069)	1.1516 (0.131)
Pension expenditure	Pension expenditure (% of GDP)	1.1006*** (0.027)	1.0953*** (0.025)	1.0524* (0.032)	1.4128*** (0.060)
GDP	GDP per capita, PPP (1000 constant 2005 international $)	1.0277*** (0.007)	1.0396*** (0.007)	1.0383*** (0.010)	1.0412*** (0.015)
Problem pressure	Population aged 65 and above (% of total)	0.9805 (0.019)	0.9493*** (0.019)	0.9535* (0.024)	0.7657*** (0.030)
Electoral cycle	Electoral cycle variable dummy (1: year of or after election, 0: remaining years)	1.0063 (0.074)	1.2469*** (0.095)	1.1089 (0.107)	1.7027*** (0.278)
Inflate					
Regulatory style	Average number of dismantling events (rolling annual mean in ongoing legislature)	0.8959*** (0.036)	0.6938*** (0.049)	0.7550*** (0.057)	1.0442 (0.039)
N		566	566	566	566
Wald chi^2		1267.24	1109.09	389.04	137.72
P		0.000	0.000	0.000	0.000

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Budge et al., 2001; Klingemann et al., 2006; Norris, 2009; Knill et al., 2010; Armingeon et al., 2011; Volkens et al., 2011; OECD, 2012; World Bank, 2012.

change by one unit. Thus, the estimated value refers to the relative change in the rate of occurrence (that is, the number of instances of policy dismantling in a given year) that is associated with a one-unit change in the explanatory variable of interest holding all other independent factors constant. The explanatory factors are lagged by one unit (that is, one year), as decisions to dismantle policies as a response to socio-economic changes are expected to require a certain period in order to be translated into actual dismantling policies.

The dependent variable is operationalized first by taking the overall sum of policy dismantling in a country per year. The second model takes the sum of intensity dismantling decisions while models three and four in Table 11.2 refer to the two subgroups of change – level and scope respectively. The different operationalizations of the dependent variable (as reported in Table 11.2) allow for the following conclusions. Except for the last column (dismantling in scope only), the analyses of the data lend confirmation to hypotheses H3 and H4 put forward in the *Concepts and theories* section with respect to the quantitative approach. Due to the incidence ratio format of the estimates, coefficients below one are indicative of a negative relationship between the independent and dependent variables, while values greater than one imply positive correlation. More concretely, the estimates suggest that an average increase in the relevant sentences of the party manifestos (as weighted by their members' seats in parliament) by 1 unit leads to a change in the incidence rate of (total) policy dismantling by a factor of 1.0837, hence an increase by 8.4 per cent. The values are even stronger from the policy intensity perspective (columns four to six) with values up to 1.236 (or 24%) with respect to the levels of pension payments. As regards the expected impact of the development in pension expenditures on the actors' willingness to engage in policy dismantling, the findings are confirmative of the hypothesized positive relationship. According to the estimates presented in Table 11.2, an increase in the share of pension payments of a country's GDP by 1 per cent leads to increases in policy dismantling by a factor ranging from 1.05 (with respect to the levels of pension payments) to 1.41 (regarding the tightening of eligibility criteria or the scope). These values imply increases in the probability of policy dismantling of 5 to 41 per cent. Altogether, hypotheses H3 and H4 regarding the impact of political preferences and financial problem pressure, respectively, are confirmed by the results.[5]

In sum, the qualitative and the quantitative analysis yield results that are of complementary value in order to deepen our understanding of the processes and factors of policy dismantling in the field of public pensions. Hypotheses H1 (qualitative approach) and H3 (quantitative

perspective) make predictions as regards the influence of veto player (H1) and actor preference (H3) structures for the likelihood of policy dismantling. The findings suggest that a high prioritization of the issue of welfare state limitation within the legislative branch (that is, parliament) constitutes an incentive for policy-makers to engage in dismantling the public pension scheme. As a consequence of a fragmented veto-player structure (that is very likely in generally welfare-friendly environments), proponents of dismantling (that is, policy entrepreneurs) appear to prefer incremental or hidden strategies to realize their plans to reduce the generosity of the scheme. The results for hypotheses H2 (qualitative) and H4 (quantitative perspective) offer a similar complementary interpretation. While the qualitative analysis of the data rejects a causal relationship between the general economic pressure or downturn and the dismantling of pension policies, the quantitative analysis also refutes the systematic correlation of the two variables. As a consequence of these findings, it is highly improbable that the economic development of a country has a direct and immediate effect on the decision-makers' willingness to dismantle the public pension scheme. This factor might, however, still have indirect effects on policy dismantling as an interacting or intervening variable. In order to refine the theoretical model in this sense, one would certainly have to extend the analysis by including and testing a broader range of explanatory factors. Finally, the testing of hypothesis H4 (quantitative) suggests a positive relationship between the financial pressure on the pension scheme and the likelihood of policy dismantling. The strong correlation of the two variables in the expected direction suggests that policy-makers are responsive to these immediate and policy-related indications of problem pressure (H4) rather than to the general development of the country's economy (H2). Altogether, the empirical findings of the two distinct analyses offer complementary and confirmatory results for the hypotheses and the underlying theoretical model. This way, we have been able to enhance the robustness of the conclusions with respect to the dynamics and factors of policy dismantling.

Conclusion: Trade-off between breadth and depth

This paper documents the application of a mixed-methods design in the field of comparative public policy research. With the help of concrete empirical data, we showed how different – qualitative and quantitative – methods can be integrated in a complementary research design in order to obtain more robust research results.

In principle, applied mixed-methods research can be divided into four distinct ways to link quantitative and qualitative approaches depending on the point of departure of the analysis and the underlying research aim (Morse, 1991; Teddlie and Tashakkori, 2006; Teddlie and Yu, 2007). First, if both research endeavours are of exploratory nature, the scientific goal might be to build extensive theoretical frameworks and to derive testable hypotheses. Second, in situations in which the small-N analysis precedes the large-N study, the qualitative approach might help identify causal mechanisms and hypotheses, which are subsequently subject to quantitative testing. Third, and more rarely, the two research steps might also be reversed with the large-N exploratory study preceding the small-N confirmatory analysis. Fourth, both methods might be applied simultaneously with the objective to extensively test a previously generated theory as well as the derived hypotheses from different perspectives. In ideal terms, this extensive deductive strategy extends the credibility of the results and estimates if the two strands of analysis arrive at similar and complementary conclusions.

In this paper, we opted for a confirmatory deductive analysis of policy dismantling in pension politics taking a comparative perspective. Given the complexity of the research topic we chose a complementary design that allows for testing theoretically derived hypotheses with the help of different methodological techniques (for example, by combining a quantitative Poisson regression with qualitative case studies). More specifically, we tried to assess two theoretical propositions from different angles by identifying causal chains on the one hand and mean causal effects on the other. The first set of theoretical propositions was centred on the political actors and their preferences (micro side) while the second set dealt with economic and financial conditions as factors to determine government decisions (macro side). Altogether, the integrated application of the two strands of methods helped us to disentangle causes and conditions of policy dismantling in the field of pension reform.

Even though mixed-methods designs help to compensate for the deficits of different single methodological approaches, one has to critically evaluate – ex-ante and ex-post – if the associated gains exceed the costs of such research endeavours. For practical and financial reasons, ambitious mixed-methods designs mostly depend on well-funded large-scale research projects. Before engaging in such a design, one has to critically ask whether the means (for example, the method) actually match the end (that is, the required financial and personnel outlays). Further, choosing the 'appropriate' mixed-methods is not a trivial exercise. In the light of the great variety of different single methods, it is even more challenging to select the appropriate combination of two methods taking into

account both the state-of-the-art of the research topic and the strengths and weaknesses of each method. In short, the designing of efficient mixed-methods research designs requires not only considerable expertise in the field of study but also deep knowledge of the research methodologies of different schools.

Notes

1 For useful practical compendia or overview articles for applied mixed-methods research see Greene et al. (2005), Bryman (2006), Creswell (2010), Plano et al. (2008), Tashakkori and Teddlie (2010).
2 In this context, see, for example, Lieberman (2005) for a useful introduction to nested analysis as a subtype of sequential mixed-methods design.
3 For a thorough discussion of the conceptual and philosophical background of mixed-methods and research paradigms, see Johnson and Onwuegbuzie (2004), Giddings (2006), Mahoney and Goertz (2006), Morgan (2007), Denscombe (2008), Denzin (2010).
4 A conceptual experiment in strategic interaction analysis and comparative statics prescribes that either the micro side of individual actors' preferences are varied, while holding constant environmental factors, such as institutional rules in order to predict an outcome, or inversely, actors' preferences are held constant and environmental factors are varied in order to predict an outcome (Lake and Powell, 1999).
5 In this context, it is important to note that the probability-based approach does not imply the testing of the hypotheses as such. Rather, the estimates allow for the rejection of the associated null hypotheses that are at the basis of the different regression models. This rejection leads to the acceptance of the alternative hypothesis for the single coefficient estimates. More concretely, the null hypothesis states that the explanatory variables and the dependent variable do not covariate systematically. The levels of significance delimit the probability of rejection of this claim. The research hypothesis is tentatively accepted in case the associated null hypothesis can be rejected on a given level of significance. The regression results give strong evidence for the rejection of the nulls of the prioritisation and the financial pressure hypotheses as the coefficients point to the expected direction with estimated levels of significance of up to 1 per cent.

Bibliography

Armingeon, Klaus, David Weisstanner, Sarah Engler, Panajotis Potolidis, Marlène Gerber and Philipp Leimgruber (2011) 'Comparative Political Data Set I (23 OECD Countries)', Retrieved 23 August 2012 from http://www.ipw.unibe.ch/content/team/klaus_armingeon/comparative_political_data_sets/index_ger.html.
Bauer, Michael W. and Christoph Knill (2012) 'Understanding Policy Dismantling: An Analytical Framework', in Michael W. Bauer, Andrew Jordan, Christoffer Green-Pedersen and Adrienne Héritier (eds) *Dismantling Public Policies: Preferences, Strategies, and Effects* (Oxford University Press), 30–55

Bonoli, Giuliano (2007) 'Switzerland: The Impact of Direct Democracy', in Ellen M. Immergut, Karen M. Anderson and Isabelle Schulze (eds) *The Handbook of West European Pension Politics* (Oxford University Press), 203–47.

Brewer, John and Albert Hunter (1989) *Multimethod Research: A Synthesis of Styles* (Newbury Park: Sage).

Bryman, Alan (2006) 'Integrating Quantitative and Qualitative Research: How Is It Done?', *Qualitative Research*, 6 (1), 97–113.

Budge, Ian, Hans-Dieter Klingemann, Andrea Volkens, Judith Bara and Eric Tanenbaum (2001) *Mapping Policy Preferences: Estimates for Parties, Electors and Governments 1945–1998* (Oxford University Press).

Campbell, Donald T. and Donald W. Fiske (1959) 'Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix', *Psychological Bulletin*, 56 (2), 81–105.

Caporaso, James A. (2009) 'Is There a Quantitative-Qualitative Divide in Comparative Politics? The Case of Process Tracing', in Todd Landman and Neil Robinson (eds) *The SAGE Handbook of Comparative Politics* (Thousand Oaks: Sage), 67–83.

Coppedge, Michael (1999) 'Thickening Thin Concepts and Theories: Combining Large N and Small in Comparative Politics', *Comparative Politics*, 31 (4), 465–76.

Corbin, Juliet M. and Anselm C. Strauss (1998) *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, 2nd edn (Thousand Oaks: Sage).

Creswell, John W. (2003) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (Thousand Oaks: Sage).

Creswell, John W. (2010) 'Mapping the Developing Landscape of Mixed Method Research', in Abbas Tashakkori and Charles Teddlie (eds) *SAGE Handbook of Mixed Methods in Social and Behavioral Research* (Thousand Oaks: Sage), 45–68.

Denscombe, Martyn (2008) 'Communities of Practice: A Research Paradigm for the Mixed Methods Approach', *Journal of Mixed Methods Research*, 2 (3), 270–83.

Denzin, Norman K. (2010) 'Moments, Mixed Methods, and Paradigm Dialogs', *Qualitative Inquiry*, 16 (6), 419–27.

Doyle, Louise, Anne-Marie Brady and Gobnait Byrne (2009) 'An Overview of Mixed Methods Research', *Journal of Research in Nursing*, 14 (2), 175–85.

Fearon, James D. and David D. Laitin (2008) 'Integrating Qualitative and Quantitative Methods', in Janet M. Box-Steffensmeier, Henry E. Brady and David Collier (eds) *The Oxford Handbook of Political Methodology* (Oxford University Press), 756–76.

Ferrera, Maurizio and Matteo Jessoula (2007) 'Italy: A Narrow Gate for Path-Shift', in Ellen M. Immergut, Karen M. Anderson and Isabelle Schulze (eds) *The Handbook of West European Pension Politics* (Oxford University Press), 396–453.

Frieden, Jeffry A. (1999) 'Actors and Preferences in International Relations', in David A. Lake and Robert Powell (eds) *Strategic Choice and International Relations* (Princeton University Press), 39–76.

Gerring, John (2005) 'Causation: A Unified Framework for the Social Sciences', *Journal of Theoretical Politics*, 17 (2), 163–98.

Gerring, John (2007) *Case Study Research: Principles and Practices* (New York: Cambridge University Press).

Giddings, Lynne S. (2006) 'Mixed-Methods Research: Positivism Dressed in Drag?', *Journal of Research in Nursing*, 11 (3), 195–203.

Greene, Jennifer C. (2008) 'Is Mixed Methods Social Inquiry a Distinctive Methodology?', *Journal of Mixed Methods Research*, 2 (1), 7–22.

Greene, Jennifer C., Holly Kreider and Ellen Mayer (2005) 'Combining Qualitative and Quantitative Methods in Social Inquiry', in Bridget Somekh and Cathy Lewin (eds) *Research Methods in the Social Sciences* (Thousand Oaks: Sage), 274–82.

Hall, Bronwyn and Kirsten Howard (2008) 'A Synergistic Approach: Conducting Mixed Methods Research with Typological and Systemic Design Considerations', *Journal of Mixed Methods Research*, 2 (3), 248–69.

Hitchcock, John H. and Bonnie K. Nastasi (2011) 'Mixed Methods for Construct Validation', in Malcom Williams and W. Paul Vogt (eds) *The SAGE Handbook of Innovation in Social Research Methods* (Thousand Oaks: Sage), 249–68.

Holland, Paul W. (1986) 'Statistics and Causal Inference', *Journal of the American Statistical Association*, 81 (396), 945–60.

Jick, Todd D. (1979) 'Mixing Qualitative and Quantitative Methods: Triangulation in Action', *Administrative Science Quarterly*, 24 (4), 602–11.

Johnson, R. Burke and Anthony J. Onwuegbuzie (2004) 'Mixed Methods Research: A Research Paradigm Whose Time Has Come', *Educational Researcher*, 33 (7), 14–26.

Johnson, R. Burke, Anthony J. Onwuegbuzie and Lisa A. Turner (2007) 'Toward a Definition of Mixed Methods Research', *Journal of Mixed Methods Research*, 1 (2), 112–33.

King, Gary, Robert O. Keohane and Sidney Verba (1994) *Designing Social Inquiry: Scientific Inference in Qualitative Research* (Princeton University Press).

Kittel, Bernhard (2006) 'A Crazy Methodology? On the Limits of Macro-Quantitative Social Science Research', *International Sociology*, 21 (5), 647–77.

Klingemann, Hans-Dieter, Andrea Volkens, Judith Bara, Ian Budge and Michael McDonald (2006) *Mapping Policy Preferences II: Estimates for Parties, Electors, and Governments in Eastern Europe, the European Union and the OECD 1990–2003* (Oxford University Press).

Knill, Christoph, Kai Schulze and Jale Tosun (2010) 'Politikwandel und seine Messung in der vergleichenden Staatstätigkeitsforschung: Konzeptionelle Probleme und mögliche Alternativen', *Politische Vierteljahresschrift*, 51 (3), 409–32.

Lake, David A. and Robert Powell (1999) *Strategic Choice and International Relations* (Princeton University Press).

Lieberman, Evan S. (2005) 'Nested Analysis as a Mixed-Method Strategy for Comparative Research', *American Political Science Review*, 99 (3), 435–52.

Mahoney, James (2008) 'Toward a Unified Theory of Causality', *Comparative Political Studies*, 41 (4–5), 412–36.

Mahoney, James and Gary Goertz (2006) 'A Tale of Two Cultures: Contrasting Quantitative and Qualitative Research', *Political Analysis*, 14 (3), 227–49.

Miles, Matthew B. and A. Michael Huberman (1994) *Qualitative Data Analysis: An Expanded Sourcebook*, 2nd edn (Thousand Oaks: Sage).

Morgan, David L. (2007) 'Paradigms Lost and Pragmatism Regained: Methodological Implications of Combining Qualitative and Quantitative Methods', *Journal of Mixed Methods Research*, 1 (1), 48–76.

Morse, Janice M. (1991) 'Approaches to Qualitative-Quantitative Methodological Triangulation', *Nursing Research*, 40 (2), 120–23.

Morse, Janice M. (2010) 'Procedures and Practice of Mixed Method Design: Maintaining Control, Rigor, and Complexity', in Abbas Tashakkori and Charles

Teddlie (eds) *SAGE Handbook of Mixed Methods in Social and Behavioral Research* (Thousand Oaks: Sage), 339–52.

Neuman, W. Lawrence (2002) *Social Research Methods: Qualitative and Quantitative Approaches*, 5th edn (Boston: Allyn and Bacon).

Norris, Pippa (2009) 'Democracy Time-Series Dataset', Retrieved 23 April 2012 from http://www.hks.harvard.edu/fs/pnorris/Data/Data.htm

OECD (2012) 'OECD Statistics', Retrieved 17 April 2012 from http://www.oecd-ilibrary.org/statistics.

Onwuegbuzie, Anthony J. and Nancy L. Leech (2006) 'Linking Research Questions to Mixed Methods Data Analysis Procedures', *The Qualitative Report*, 11 (3), 474–98.

Plano Clark, Vicki L., John W. Creswell, Denise O'Neil Green and Ronald J. Shope (2008) 'Mixing Quantitative and Qualitative Approaches: An Introduction to Emergent Mixed Methods Research', in Sharlene Nagy Hesse-Biber and Patricia Leavy (eds) *Handbook of Emergent Methods* (New York/London: The Guilford Press), 363–87.

Ragin, Charles C. (1989) *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies* (Berkeley: University of California Press).

Schmitt, Sophie (2012) 'Sequences of Active Policy Dismantling? Path Dependency in Pension Reform Processes', in Michael W. Bauer, Andrew Jordan, Christoffer Green-Pedersen and Adrienne Héritier (eds) *Dismantling Public Policies: Preferences, Strategies, and Effects* (Oxford University Press), 57–80

Seawright, Jason and John Gerring (2008) 'Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options', *Political Research Quarterly*, 61 (2), 294–308.

Sobel, Michael E. (2000) 'Causal Inference in the Social Sciences', *Journal of the American Statistical Association*, 95 (450), 647–51.

Tarrow, Sidney (2004) 'Bridging the Quantitative-Qualitative Divide', in Henry E. Brady and David Collier (eds) *Rethinking Social Inquiry: Diverse Tools, Shared Standards* (Lanham: Rowman and Littlefield), 171–79.

Tashakkori, Abbas and Charles Teddlie (2010) *SAGE Handbook of Mixed Methods in Social and Behavioral Research* (Thousand Oaks: Sage).

Teddlie, Charles and Abbas Tashakkori (2006) 'A General Typology of Research Designs Featuring Mixed Methods', *Research in the Schools*, Special Issue, 13 (1), 12–28.

Teddlie, Charles and Fen Yu (2007) 'Mixed Methods Sampling: A Typology with Examples', *Journal of Mixed Methods Research*, 1 (1), 77–100.

Volkens, Andrea, Onawa Lacewell, Pola Lehmann, Sven Regel, Henrike Schultze and Annika Werner (2011) *The Manifesto Data Collection: Manifesto Project (MRG/CMP/MARPOR)* (Berlin: Wissenschaftszentrum Berlin für Sozialforschung).

World Bank (2012) 'World Development Indicators (WDI)'. Retrieved 8 July 2012 from http://data.worldbank.org/data-catalog/world-development-indicators.